AN ABSTRACT OF THE DISSERTATION OF

Heather E. Lintz for the degree of Doctor of Philosophy in Botany and Plant Pathology presented on December 7, 2010.

Title: Ecological Thresholds, Climate Extremes, and Tree Species' Distributions across the Pacific Coastal United States.

Abstract approved:

_____

Bruce McCune

Species' distributions across the landscape are perhaps the least understood yet most conspicuous features of life on earth. Ecologists have long studied species' distributions; yet, many questions remain about why species occur where they do. Such questions persist largely because species' distributions are complex systems with challenging properties like non-linearity, high dimensionality, and strong interactions. One impediment to our understanding of species' distributions is the use of standard quantitative approaches, which have limitations when applied to data from complex systems. The objective of this work is to develop tools that can be used to improve our understanding of natural complexity, in particular, factors relating to species' distributions at large scales. Our study systems are tree species' distribution from Forest Inventory Analysis (FIA) sites across the Pacific coastal USA.

Ecological thresholds are an example of complex system behavior. While an ecological threshold is a widely accepted concept, most empirical methods have not quantified the direct drivers of thresholds. Causal understanding of thresholds detected

empirically requires their investigation in a multi-factor domain containing the direct drivers (often referred to as state space). This work develops an approach to quantify thresholds from response surfaces in three-dimensional state space. Two new indices of shape attributes are measured from response surfaces: threshold strength and diagonality.

Our ability to describe and probe the basis of species' distributions depends on the quality of our data. The most extensive and thorough field data set on tree species in North America is provided by the FIA program. Because the FIA inventory recently changed from an amalgam of different sampling approaches to a nationally-standardized approach in 2000, the two types of inventories represent different probabilities of detecting trees per sample unit. The application of non-parametric multiplicative regression to build and compare niche models for 41 tree species from the old and new FIA design shows two likely effects of differences in inventory approach on niche models and their predictions. First, there is an increase from 4 to 6% in random error; this increase is noted for modeled predictions from the different inventories when compared to modeled predictions from two samples of the same inventory. Second, systematic error (or directional disagreement among modeled predictions) is detectable for 4 out of 41 species among the different inventories.

Models of tree species' distributions rarely incorporate historical extremes in climate as predictors. I examine this by pitting climate means versus climate extremes to determine species probability of occurrence by life stage. Interactions between climatic oscillations are characterized in a new way to define climate extremes and make them computationally tractable as predictors of species' distributions. Results indicate that 27% of the models across 22 species and seven climate variables show a climate extreme explaining more variability than the climate mean. Extremes associated with freeze-thaw events, seasonality of precipitation, and winter minimum temperature are most frequently represented.

Overall, this study contributes statistical approaches tailored for use with complex data and provides insights into species' distributions of trees across the Pacific coastal USA. In particular, this study advances the detection and measurement of thresholds in

response surfaces, the evaluation of the consequences of a major change in sampling and measurement methodology, and the relationship of both climatic extremes and means to species' distributions.

Ecological Thresholds, Climate Extremes, and Tree Species' Distributions across the Pacific Coastal United States

by

Heather E. Lintz

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented December 7, 2010
Commencement June 2011

Doctor of Philosophy dissertation of Heather E. Lintz presented on December 7, 2010

APPROVED:

_____

Major Professor, representing Botany and Plant Pathology

_____

Chair of the Department of Botany and Plant Pathology

_____

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

_____

Heather E. Lintz, Author

ACKNOWLEDGEMENTS

I am grateful to Bruce McCune for excellent mentorship. Bruce provided the freedom to develop and pursue my own ideas while giving an ideal amount of guidance, support, and constructive criticism. Through persistent idea-sparking, questioning, and encouragement, he honed my skill in scientific thinking. He also gave me a necessary ingredient for success, confidence. Thanks to Bruce, I know how to follow my gut when it comes to contributing creatively to my field of study. His impact on my education will last my lifetime. I could not imagine a better PhD experience because of his mentorship.

I also extend great thanks to my committee members, Andrew Gray, Alix Gitelman, Ron Neilson, and Steve Sharrow. My dissertation would not have been possible without their support, patience, and feedback. In particular, I want to thank Andrew for his keen comments, ideas, and active participation in the development of this work.

Finally, I would like to thank my family. Thanks mom for cultivating my creativity and giving me the confidence to achieve. Thanks dad for teaching me strength and endurance in the face of challenge. Thanks Sharon for being a rock and so brilliant and hilarious at all hours. A super special thank you goes to my daughter, Zeah Frederick, for being the most amazing and enriching sprite, and for putting up with my long work hours. A very heartfelt thank you is for my partner, Jeff, for his unfaltering and loving support and fun companionship. A last-but-not-least thank you extends to the whole McCune-Muir family for being generous, nurturing, and fabulous role models for Zeah.

CONTRIBUTION OF AUTHORS

Dr. Bruce McCune contributed to the entire dissertation through guidance and editing. Bruce also provided support for implementation of Non-parametric Multiplicative Regression and Multivariate Response Permutation Procedure in the software HyperNiche and PC-ORD. Dr. Andrew Gray contributed to every chapter through his expertise with the Forest Inventory and Analysis database, data extraction and documentation, guidance, and editing. Dr. Kate McCulloh contributed a small portion of data and one paragraph of text in Chapter 1.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

Ecological Thresholds, Climate Extremes, and Tree Species' Distributions across the Pacific Coastal United States

Chapter 1. Introduction

"Progress in science depends on new techniques, new discoveries and new ideas, probably in that order."
Sydney Brenner, 2002

The field of ecology addresses the most complex natural systems on Earth. Ecological systems exhibit complex interactions, hierarchical relations, and emergent non-linear behavior like thresholds (Holling 1973, Allen and Starr 1998, Jørgensen et al. 1992, Maurer 1998). The study of complex systems bridges mathematics, statistics, computer science, and disciplines dealing with complex system behavior like ecology (Érdi 2008). Although there is a strong sub-discipline of ecology devoted to systems study (e.g. Jørgensen et al. 1992, Odum 1971), and another to statistical methodology (e.g. Elith et al. 2006, Oksanen and Minchin 2002), many aspects of the science of complex systems have yet to find home in ecological thought and methodology. Thus, it is not surprising that identification of underlying theory in ecology is still challenged by analytical constraints (Weng et al. 1999, Taylor 2005). Development of quantitative tools that are tractable with principles from complex systems are needed to make headway with complex ecological phenomena. Computer intensive statistical methods or data mining approaches are a starting point for statistical questions that embrace complexity (Efron and Tibshirani 1991). Examples of questions that incorporate ecological complexity include: what drives species' occurrence on the landscape, and how will species' occurrence respond to climate change?

To date, understanding of species' distributions is based on an important concept in ecology: the species' niche. The relationship between a species and its habitat is part of an '*n*-dimensional hypervolume' or species' niche that describes conditions where a

species can persist (Hutchinson 1957). The niche concept follows a long history in ecology (e.g. Grinnell 1917, Gause 1934, Hutchinson 1957, MacArthur 1972, Austin et al. 1990). Contemporary niche models have widespread and diverse use including: species' conservation (e.g. Hannah et al. 2002, Marini et al. 2009), species' re-introduction (Yanez and Floater 2000), species' migration and invasion (e.g. Woodall et al. 2009b, Crossman and Cooke 2011, Peterson 2003, Thuiller et al. 2004), biodiversity conservation (e.g. Thuiller et al. 2005, Crossman and Cooke 2011), specimen collection (e.g. Jarvis et al. 2005, Raxworthy et al. 2003), the discovery of new species (Raxworthy et al. 2003), and predicting species' response to climate change (e.g. Peterson et al. 2002, Shafer et al. 2001, Rehfeldt et al. 2006). Niche models are also used for basic research that targets causal understanding related to topics such as speciation (e.g. Graham et al. 2004; Peterson et al. 1999), species' co-occurrence (e.g. Kelly et al. 2008), species' distributions (e.g. Engelbrecht et al. 2007, Svenning and Skov 2004), and molecular diversity (e.g. Hugall et al. 2002).

Most niche models rely on statistical relationships. Apart from the works described above, a smaller subset of approaches merge different statistical relationships (often at a finer physiological scale) using equations that are applied sequentially in time (e.g. Chuine and Beaubien 2001, Landsberg and Waring 1997). These are known in ecology as dynamic models and they are used for various purposes including forecasting. There are varying opinions over the relative advantages and pitfalls of dynamic modeling versus purely empirical modeling. The bottom line is that both approaches are needed to make headway with the complexity of species niches, and there is ample room for improvement in both approaches. This thesis examines top-down, empirical statistics in niche modeling. It focuses on improving our understanding of natural complexity and species' distributions by improving our toolsets to evaluate emergent pattern at the scale it is produced.

The need for methodological improvements in niche modeling has been identified over the past decade (e.g. Elith et al. 2006, Guisan and Zimmerman 2000, Araújo and Guisan et al. 2006, Hampe 2004, Austin 2002). Chapters herein tackle a subset of

methodological issues from the viewpoint that challenges associated with method development come from the complexity of nature and its related statistical properties. These properties can manifest themselves in the shape of a response surface. These properties can also manifest themselves in a more complex way through interactions occurring at different scales in space and time. One major challenge for this thesis is that adequate representation of response surfaces using a statistical model has not been widely addressed for a species' niche. The main questions driving my work are thus: How can we better characterize emergent behavior and interactions coming from complex systems through statistical modeling? How can we better capture the "true" shape of an empirical response given the potential complexity of underlying interactions and resulting non-linearity? How do current tools stack up in this regard? How can we characterize complex interactions occurring in space and time that affect a species' niche? How might new analytical approaches open doors to advancing ecological theory? I believe answers to these questions can change the trajectory of our discipline to one more directed toward complex systems research. Answers to these questions can also increase our understanding of emergent ecological phenomena and help move ecology toward a more predictive science. In addition to addressing the above questions, this dissertation examines a related issue associated with data integration for the species data that I use in each chapter, data from the U.S.D.A. Forest Inventory and Analysis Program.

Chapter 2 develops a method to quantify thresholds with respect to more than one driver in state space (or the space containing thresholds responses and their direct drivers). Prior to this research, no method was available to quantify thresholds with respect to more than a single driver. The first step involves testing different data mining methods with respect to how well they capture the shape of the underlying data structure. This serves as preliminary work toward developing two new indices. The indices demonstrate how to answer mechanistic questions related to ecological thresholds using emergent properties of response surfaces, namely shape attributes.

Chapter 3 examines the statistical effect of a massive overhaul in a forest inventory. The overhaul occurred in the year 2000 for the Forest Inventory and Analysis

program (FIA). Data from FIA represent the largest data set of *in situ* forest monitoring in North America. Data on tree presence/absence date back several decades. Although the data are ripe for trend detection, the change in inventory type can confound trends among measurements made before and after the year 2000. Here, I ask whether the change in inventory affected niche models for tree species across the western United States. A new method in niche modeling was used for this work, Non-parametric Multiplicative Regression (NPMR) (McCune 2006). NPMR represents the nature of organismal response to multiple interacting factors (McCune 2006). This method was also tested in Chapter 2.

Chapter 4 explores if there is current evidence to support climate extremes explaining more variability than means in species' probability of occurrence. I also ask if the answer depends on life stage for tree species. The question is based on the zeitgeist in ecological theory with respect to species' niches surging from a recent blend of paleo-climate and paleo-ecological works (e.g. Jackson et al. 2009). Juvenile niches for tree species are thought to be narrower than adults and thus more sensitive to climate extremes. This is hypothesized to enhance episodic recruitment or pulses of tree establishment coinciding with particular temporal windows of climatic opportunity. The use of unconventional methodology to identify and characterize complex interactions in climate plays an important role in gaining access to the question. NPMR is also used. Finally, I develop two new indices to measure climate climatic phenomena important to plants, namely spring and fall freeze-thaw events. Such events occur during vulnerable periods of hardening and de-hardening for plants and can cause freeze damage and mortality.

In summary, each of the chapters in this thesis map to my overall objective to forge new paths toward a better acknowledgement and characterization of complexity in nature through method development. The results of this work and directions for future study are summarized in Chapter 5.

Chapter 2. Quantifying Ecological Thresholds from Response Surfaces

Heather E. Lintz, Bruce McCune, Andrew N. Gray, Katherine A. McCulloh

Ecological Modelling
Elsevier, 201 Mission Street, 26th Floor, San Francisco, CA 94105
In press, doi:10.1016/j.ecolmodel.2010.10.01

ABSTRACT

Ecological thresholds are abrupt changes of ecological state. While an ecological threshold is a widely accepted concept, most empirical methods detect them in time or across geographic space. Although useful, these approaches do not quantify the direct drivers of threshold response. Causal understanding of thresholds detected empirically requires their investigation in a multi-factor domain containing the direct drivers (often referred to as state space). Here, we present an approach to quantify thresholds from response surfaces modeled empirically in state space. We present two indices of shape attributes measured from response surfaces. The response surfaces are built using a regression method in state space. The indices are threshold strength (*T*) and diagonality (*D*). We use 48 simulated response surfaces of different shapes to test the efficacy of the indices in 3-D. Our results show that *T* is sensitive to the steepness of the transition from one state to the next, with various forms of abrupt, centralized thresholds yielding the highest values among the simulated surfaces. *D* represents the orientation of the response surface or the simultaneous influence of more than one predictor in eliciting the response gradient. Strongly diagonal surfaces have the most diagonal surface area demonstrated by sharply undulating diagonal surfaces. Given that the success of *T* and *D* requires a regression method to accurately capture any shape of complex data structure, we also test the accuracy of empirical regression methods known to be tractable with complex data. We test Classification and Regression Trees (CART), Random Forest, and Non-Parametric Multiplicative Regression (NPMR) for binary and continuous responses. We use the 48 simulated response surfaces to test the methods, and we find that prediction accuracy depends on both the *T* and *D* of the simulated data for each method. We choose the most accurate method among those we test for capturing any shape of response surface from real data, NPMR. Finally, we use NPMR to build response surfaces and quantify *T* and *D* from real ecological data sets. We demonstrate how measuring threshold strength and diagonality from multi-factor response surfaces can advance ecology.

INTRODUCTION

Ecological thresholds are an increasing research priority among natural, earth, and social sciences (USCCP 2009, Andersen et al. 2009). Simply defined, ecological thresholds are a non-linear response where a small change in the input produces an abrupt change in the output for the scale at hand (USCCP 2009, Groffman et al. 2006, Andersen et al. 2009). The occurrence of ecological thresholds can carry profound societal risks especially in the face of unprecedented environmental change (USCCP 2009). Examples of ecological thresholds include shifts in water clarity of lakes caused from continuous nutrient loading that passes a critical point (Scheffer et al. 1993) and the conversion of arctic tundra to shrubland triggered by a slight increase in temperature (USCCP 2009). Such threshold behavior is common across diverse systems and scales and represents adaptive, complex behavior (Levin 1999, Holling 1992).

Despite their importance, the mathematical characterization of ecological thresholds is poorly developed. Current methods that quantify thresholds focus either on threshold or change-point detection in time (Andersen et al. 2009) or across geographic space (Fortin 1994, Jacquez et al. 2000). Yet, thresholds can be represented in state space, geographic space, or time. While thresholds may be observed in time and geographic space, the causal drivers of thresholds are found in state space (Scheffer and Carpenter 2003).

Surprisingly, few methods exist for the quantification of ecological thresholds in state space, and for those that do, most detect the location of the threshold and apply only to a single predictor (e.g. Baker and King 2009, Brenden et al. 2008, Damgaard 2006, Toms and Lesperance 2003). In fact, to our knowledge, no method addresses the challenges that arise in state space when thresholds are characterized with respect to more than one predictor. Given the inherent complexity of ecosystems, empirical characterization of thresholds with respect to more than one predictor (or driver) is clearly warranted (Limburg et al. 2002). The most important reason for expanding threshold analysis to multi-factor predictor domains is simple. In higher dimensional

predictor space, one can detect and measure thresholds that would not be observed by analyses limited to single predictors.

One grand challenge of measuring thresholds in higher dimensional state space is that ecological thresholds can take more than one geometric form. Ecological thresholds (along with other shapes that emerge from complex systems) can result from complex behavior including interactions, hierarchical relationships, and other forms of nonlinearity (Goldenfield and Kadanoff 1999, Weng et al. 1999, Limburg et al. 2002, Kinzig et al. 2006, Andersen et al. 2009). Such complex behavior can yield many different response shapes. For example, thresholds in 3-D can be oriented perpendicular or diagonal with respect to the input gradients, they can look like Niagara Falls, or they can be confined to part of a response surface. Thus, the quantitative assessment of thresholds in *n*-dimensional state space is not as simple as fitting parametric equations, such as the logistic curve, to data. Parametric regression equations yield a distinct geometric shape or type of shape (e.g. planes or logistic curves depending on the class of equation). Consequently, by its nature, parametric regression imposes specific shapes or shape families *a priori* on data patterns. However, in complex data analysis, prudence calls for regression methods that can easily adapt to any response shape. The shape of a response surface is an emergent property of the underlying system. It warrants accurate capture, quantitative assessment, and interpretation. Unless a specific shape is expected or of interest, it should be treated as unknown prior to exploratory analysis, and ideally, exploratory analysis would use a method that does not impose a specific shape *a priori*.

An ecological threshold can be considered a type of response shape, and non-parametric regression may be the best option for assessment of multi-factor shapes or thresholds in state space. Our use of the term ‘non-parametric regression’ follows the definition for ‘computer-intensive’ regression established by Efron and Tibshirani (1991) with CART and kernel smoothers as examples. Such methods are known to be tractable with complex data and rely on computationally intensive algorithms that can involve iteration and re-sampling. Non-parametric regression may avoid imposing shape-related constraints on data patterns, however, little work tests their accuracy in recovering

different shapes of response patterns. Hence, we test the prediction accuracy of non-parametric methods, particularly, we test how well they predict the true underlying shape of the data pattern. The results of this test provide us with a regression method we can use to measure shape attributes of predicted response surfaces.

Our over-arching goal is to measure the strength and orientation of multi-factor thresholds in state space. In so doing, we provide a method to verify claims of ecological thresholds and increase our understanding of the multi-factor nature of thresholds. Our method follows two general steps. First, we model a data set and generate a predictive response surface. Then, we quantify shape attributes from that surface. We are not aware of any work that quantifies shape attributes from multi-factor response surfaces as we define them.

We define threshold strength (*T*) as the abruptness of an ecological threshold in state space. We complement this index by measuring the orientation of thresholds with more than one predictor, something we call diagonality (*D*). Diagonality occurs in 3-D responses including thresholds, and its mathematical basis merits attention in the study and interpretation of response surfaces in general. Diagonality gauges the degree to which a threshold (or any other response shape) is influenced by more than one predictor. Diagonality can assist in identifying and describing complex interactions.

The specific research objectives of this paper are: to design indices of threshold strength and diagonality and validate them using numerous simulated data sets of different shape, to test the ability of nonparametric regression methods to recover a wide range of shapes of response structures or surfaces (including thresholds) from simulated data sets to optimize measurement of thresholds, and to provide examples of how measuring threshold strength and diagonality from real response surfaces can advance ecology.

METHODS

Index of threshold strength. We describe our index of threshold strength for three-dimensional response surfaces in state space. We define a response surface as a

uniform grid of predicted values generated using a model with continuous variables as input (Fig. 1). The response value is named z, while the two predictors are x and y. We also describe a two-dimensional version (see Appendix A). The central premise of the index is based on two criteria. First, the strongest thresholds have the greatest bimodality in their frequency distribution. Second, the strongest thresholds also have the greatest monotonicity (or least change in the sign of slopes across the response surface). The second criterion is designed to rule out pathological surfaces exhibiting high bimodality but showing a spatial arrangement of response values dissimilar to a threshold.

To calculate the index, the response values are divided by their maximum range to standardize among response surfaces with different ranges. We measure departure from monotonicity incrementally across the surface using a moving circular window, which we refer to as a 'spider' (Fig. 1A). Our definition of monotonicity comes from calculus, which specifies one-dimensional input. We extend the concept of monotonicity to three dimensional response surfaces by calculating the average departure from monotonicity among repeated sets of three points as we further describe. Each set represents one-dimensional input.

To measure monotonicity from a surface of points, we use a spider comprising nine adjacent points on a grid of 100 by 100 increments or 101 by 101 points. This is a fine enough grid to capture abrupt changes in slope on a response surface. Four pairs of opposing vectors sharing a center point are defined per spider, *NESW*, *NWSE*, *NS*, and *EW* (Fig. 1A). A case definition follows for each vector pair: if the two endpoints are either both above or both below the center point, then departure from monotonicity occurs, if not, then departure from monotonicity is zero. For cases expressing departure from monotonicity, the degree of the departure follows

$$\begin{aligned} NWSE_{i,j} &= \min\left\{\left|z^{*}_{i,j+2} - z_{i+1,j+1}\right|, \left|z^{*}_{i+2,j} - z_{i+1,j+1}\right|\right\} \\ NESW_{i,j} &= \min\left\{\left|z^{*}_{i+2,j+2} - z_{i+1,j+1}\right|, \left|z^{*}_{i,j} - z_{i+1,j+1}\right|\right\} \\ NS_{i,j} &= \min\left\{\left|z_{i+1,j+2} - z_{i+1,j+1}\right|, \left|z_{i+1,j} - z_{i+1,j+1}\right|\right\} \\ EW_{i,j} &= \min\left\{\left|z_{i,j+1} - z_{i+1,j+1}\right|, \left|z_{i+2,j+1}\right|\right\} \end{aligned} \tag{1-4}$$

where *z* denotes a response point within a spider, and *i* and *j* index the point on a uniform grid. To give the spider a circular footprint and approximate invariance to rotation, diagonally-oriented vectors are shortened through interpolation and interpolated points are denoted as z^* (Fig. 1A). See Appendix A for a description of the interpolation method. Departure from monotonicity for a spider is the sum of the departures, $S_{i,j} = NWSE_{i,j} + NESW_{i,j} + NS_{i,j} + EW_{i,j}$. The sum of $S_{i,j}$ across all spiders yields overall departure from monotonicity for a response surface

$$\sum_{i=1}^{n-2}\sum_{j=1}^{n-2} S_{i,j} = \sum_{i=1}^{n-2}\sum_{j=1}^{n-2} (NWSE_{i,j} + NESW_{i,j} + NS_{i,j} + EW_{i,j}) \quad (5)$$

where *n* is the number of points within one predictor dimension. $\sum_{i=1}^{n-2}\sum_{j=1}^{n-2} S_{i,j}$ is divided by the total number of paired, opposing vectors for the surface (four times the total number of spiders evaluated or $4(n-2)^2$ in three dimensions) to yield average departure from monotonicity for the surface, *K*. We calculate monotonicity (*M*) using a negative exponential function of *K*, specifically,

$$M = e^{-950K}. \quad (6)$$

M has a *y*-intercept of 1 for perfect monotonicity and an asymptote at zero for strong departures from monotonicity. We set the exponential coefficient to 950 to ensure that the low end of the range in *M* across 48 test surfaces (presented in Figure 2) approaches zero for the two most undulating test surfaces. The rank order of monotonicity of the test surfaces in Figure 2 according to *M* are virtually the same across three orders of magnitude of exponential coefficients that adequately detect departures from montonicity. Threshold strength (*T*) is the product of monotonicity, *M*, and the bimodality of the response (Eq. 8). The standard deviation (σ_z) measures the bimodality of the frequency distribution of the response where *N* is the total number of response points. The denominator is *N* instead of *N*-1 as we use the standard deviation to describe shape rather than a population sample. The standard deviation (Eq. 7) is doubled to range from 0 to 1 (Eq. 8). Threshold strength is simply a function of bimodality for perfectly monotonic surfaces or when monotonicity (*M*) is equal to one.

$$\sigma_z = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(z_i - \bar{z})^2} \ ; \tag{7}$$

$$T = 2\sigma_z M \ . \tag{8}$$

Index of diagonality. We define diagonality as how oblique or diagonal the gradient of a response surface is oriented relative to at least two predictor gradients. Diagonality represents the simultaneous influence of more than one predictor gradient in eliciting the response (Fig. 3A). For example, perfectly diagonal surface area represents equivalence among partial first derivatives for planes (e.g. the right-most plane of Figure 3A). In contrast, non-diagonal planes vary strictly with one predictor (e.g. the left-most plane of Figure 3A). Further, surfaces with traditional, statistical interactions create curvature and thus some diagonality (for example, regression models containing multiplicative terms in an additive model) (Fig. 3B), but diagonal surfaces need not have statistical interactions (e.g. the right-most plane of Figure 3A). Statistical interactions occur when the effect of one predictor on a response depends on values of another predictor or predictors.

Diagonality (*D*) is calculated for a three-dimensional response surface formed by a grid of 101 by 101 points. The grid is comprised of many four-sided polygons each defined by a unique set of four adjacent points (Fig. 1B). The vertical distance between diagonally opposed points is calculated for each polygon, and the absolute difference between these two vertical distances is termed d_g, where *g* indexes a single polygon

$$d_g = \left\| z_{i,j+1} - z_{i+1,j} \right| - \left| z_{i,j} - z_{i+1,j+1} \right\| . \tag{9}$$

Pure diagonality (*P*) is the sum of d_g across the total number of polygons (*q*). *P* is divided by the standard deviation of the response (σ_z from Eq. 7) for comparison among disparate surfaces to yield *H* or standardized pure diagonality. *H* increases linearly with the square root of *q*; hence, it is divided by the square root of *q* to yield diagonality, *D*, a variable insensitive to *q*,

$$D = \frac{H}{\sqrt{q}} \ . \tag{10}$$

Simulated data. We test the indices with simulated data sets representing varying degrees of threshold strength and diagonality common to ecological data in state space (Table 1, Fig. 2). Our choice of simulated data emerges from theoretical expectations of ecological response surfaces (e.g. Scheffer and Carpenter 2003, Austin 2007), published examples where shapes are unconstrained by modeling methods (e.g. Waring and Major 1964, Makarewicz and Likens 1975, Bartlein et al. 1986), and author experience with hundreds of ecological response surfaces. Also, several data sets are included to expand the diagonality gradient (e.g. Z46, Z47, and Z48 in Fig. 2). Although several data sets appear quite similar (Z22, Z23, and Z24 in Fig.2), they have subtle yet important differences in steepness and step height.

Method Comparison. We test the performance of each of three methods in modeling 48 simulated data sets as continuous and binary data. We select different classes of non-parametric regression methods known to be tractable with complex data (Efron and Tibshirani 1991): Classification and Regression Trees (CART) (Breiman et al. 1984), Non-parametric Multiplicative Regression (NPMR) (a kernel smoother) (McCune 2006), and a statistical ensemble method using CART as a building block, Random Forest (Breiman 2001). For each method we use the same settings across all test surfaces. We establish settings from recommendations and examples in peer-reviewed literature (explanations of methods and settings are described in Appendix A).

We compare the prediction accuracy (henceforth referred to as accuracy) of the methods by examining prediction error across all simulated shapes for binary and continuous responses. The accuracy for a continuous response is assessed with R^2. For accuracy in binary classification, we use the area under the receiver operator characteristic curve (AUC) (see Appendix A) (Hanley and McNeil 1982). Fig. 4 depicts scatterplots of accuracy versus threshold strength and diagonality for CART, Random Forest, and NPMR. Each point represents a median, externally-validated accuracy of 100 models built from random samples (N=250), which are drawn from a simulated data set (100 increments squared or size N=10,201) of specific shape; we choose N=250 as a realistic size for an ecological data set. External validation gauges prediction error for

external data. We rely on variable selection and overfitting controls inherent to each method when supplied with the two predictors (x, y).

RESULTS

Efficacy of the threshold strength index. The rank order of threshold strength is sensitive to steepness or how closely the response surfaces resemble a single step with highly undulating surfaces yielding the lowest threshold strength ($T = 0$), progressing through the Gaussian hill ($T = 0.36$) to the Gaussian ridge ($T = 0.68$) to end with various forms of strong, centralized thresholds ($T > 0.93$) (Table 1, Fig. 2). All surfaces with morphologies resembling single steps rank higher than the other shapes presented. The index tracks incremental changes in steepness among similarly-shaped monotonic surfaces such as single steps (e.g. Table 1, Z1 and Z2 in Fig.2; see Appendix A); however, increased departure from monotonicity can slightly increase with increased steepness in 'staircases' (Z24 > Z23 in Fig. 2); yet, the effect of this is not detectable at two decimal places. The index ranks surfaces resembling centralized steps similarly regardless of exact form. Thus, a central threshold showing a steep transition albeit with more curvature (from a bird's eye view) (e.g. Z3) ranks closely with a central step showing a steep transition but no curvature (e.g. Z34) (Table 1, Fig. 2). The general shape of a threshold (albeit with variable steepness) is lost below $T = 0.72$ for the sample of 48 shapes we provide (Table 1, Figure 2). Additionally, surfaces in two and three dimensions generated from the same function yield equivalence in threshold strength (see Appendix A). Finally, the index detects abrupt changes between planar features that are parallel to the *x*-*y* plane. For example, the surface Z35 contains an abrupt change between different regions of the response surface where one side of the transition is $z = 0$ (or a static value for the response variable shown as a single color, black) but the other side of the transition resembles a skate ramp (shown with the color gradient; Fig. 2). Consequently, Z35 yields a relatively low value of threshold strength (T=0.53). Although an abrupt transition exists in this surface, the transition does not contribute to a step-like

form where each state is flat and parallel to the *x-y* plane, which is the operational definition of a threshold we present here.

Efficacy of the diagonality index. The simulated 3D surfaces varying with only one predictor have diagonality of zero as expected (Table 1). Strongly diagonal surfaces have the most diagonal surface area demonstrated by sharply undulating diagonal surfaces such as 'weaving' (*D*=20.31) (Z48 in Fig. 2). The index is insensitive to scale for scales small enough to capture global shape starting with 100 increments or 101 points by 101 points for a square grid. Also, diagonality varies linearly with angle of rotation for a surface as expected (see Appendix A). However, the index does not explicitly discern the spatial location and configuration of diagonality present within a surface. For example, two different shapes of surfaces, one planar (Z17, Fig. 2) and another kite-like (Z27, Fig. 2) have very similar values of diagonality (D=2.55, and D=2.52 respectively).

Accuracy of modeling methods with simulated data. The accuracy of each method depends on the threshold strength and diagonality of the original data structure with each method differing in degree of dependence (Fig. 4). The accuracy of most methods decreases as diagonality increases and threshold strength decreases with the exception of NPMR with continuous data (lower right two axes, Fig. 4). NPMR demonstrates the least variability (seen as quantile bars in Fig. 4) and the greatest accuracy (seen as medians in Fig. 4) compared to the other methods for a given response shape. The sensitivities of modeling methods to shape attributes of data structure arises from features specific to each modeling method, which manifest in visual differences of predicted surfaces for different shapes (Fig. 5). For our subsequent analyses using real ecological data, we choose the most accurate and robust method we test, NPMR. We encourage testing of other methods.

APPLICATION OF THE INDICES

Application of threshold strength and diagonality with real data can test theory and answer questions about ecological thresholds. We present examples using real data with the goal of demonstrating how the indices can be applied (Fig. 6). The results

provided by the examples are preliminary and require further investigation. Our examples focus on thresholds in state space. However, we recognize that these tools can apply to thresholds in time and geographic space, and these are topics of future study.

For our first example, the indices evaluate the theory formulated by Berryman (1982) and reviewed by Christiansen et al. (1987) (Fig. 6A). The theory holds tree vigor and bark beetle attack as drivers of threshold responses in tree or stand survival across species. Here we evaluate the question, do bark beetle densities and tree vigor drive threshold responses in sapwood survival across tree species? Figure 6A demonstrates that the response surface of *Picea abies* survival has a moderately strong threshold (T= 0.76), while *Pinus contorta* has a weaker threshold (T=0.61). For a benchmark comparisons, see surface Z2, Fig. 2, Table 1, also with T=0.76, and a diagonally tilted plane, Z19, with T=0.41. The results suggest that the theory does apply equally well to both species for the variables tested. Also, responses of both species show diagonality; thus, each surface demonstrates that both factors elicit the response gradient among species. However, *P. abies* shows greater diagonality compared to *P. contorta* (Fig. 6A). Other factors likely need to be given account as recent works support cross-scale drivers behind bark beetle thresholds (e.g. Raffa et al. 2005, Raffa et al. 2008).

Figure 6B demonstrates an application of threshold strength in a 2-D context. Since this is in 2-D, only threshold strength can be measured. For this example we ask: does greater stomatal control (termed isohydry) create stronger thresholds in percent loss of conductivity versus water potential for woody vascular plants? Isohydric plants close their stomata (cells controlling gas exchange from leaves) when leaf water potentials reach a set value. Anisohydric plants allow water potential to decline with water stress (Vogt 2001). Vulnerability curves measure the percent loss of hydraulic conductivity of xylem (water-conducting tissue) with declining water potential; they also assess the function of water-transporting conduits within the plant during drought stress (Sperry et al. 1988) (Fig. 6B). Isohydry may create stronger thresholds in vulnerability curves of vascular plants as the strategy precludes the need for plants to construct conduits with differing resistances to water stress. The measurement of threshold strength from

vulnerability curves of iso- or anisohydric species is necessary to evaluate the research question. Preliminary calculations suggest that increased stomatal control may create stronger thresholds in vulnerability curves (T = 0.82, 0.74 for isohydric species; T = 0.62, 0.58 for anisohydric species) (Fig. 6B). However, a larger sample size consisting of more species is needed to ascertain this.

Last, threshold strength and diagonality can be applied to selected domains within a complex response surface. For example, Figure 6C shows a cropped portion of a response surface for a model of the probability of tree species' occurrence relative to climate for *Pinus ponderosa* in Oregon. The model is based on presence/absence data (Azuma et al. 2002, 2004). We select and crop the response within a specific climate domain. At first glance, one might assume that the selected portion of the response surface resembles a threshold; however, when compared to simulated data, the threshold strength is weak (T=0.68). Further, the low diagonality shows that the response within this domain is mainly driven by a single variable (D=0.76). However, the diagonality of the surface as a whole demonstrates that both drivers are responsible for eliciting the response gradient (mostly in regions outside the selected domain) (D=5.41). The lack of diagonality within the cropped domain in Figure 6C elicits the following question: why is the probability of tree species' occurrence only attributable to PCA1 within the selected domain? Response surfaces are snapshots of complex system behavior, and quantifying the diagonality (and threshold strength) of selected regions of response surfaces can identify interactions within the surfaces.

DISCUSSION

Ecological relevance. Threshold strength and diagonality represent the first tools to quantify multi-factor ecological thresholds in state space. The examples with real data demonstrate utility of the indices in state space. For example, we measure threshold strength for a diagonal response in a multi-factor state space (e.g. Figure 6A, left panel). We detected a relatively strong threshold. If this response data were to be analyzed with

respect to either one of those predictors alone, the threshold strength would be much lower.

The indices can be measured from a cropped portion of a response surface. This is a fundamental step toward using the indices within a roving window to measure *T* and *D* at different scales within the surface (Figure 6C). This can serve various research purposes such as finding the regions of strong behavior in a multi-factor response surface. Finally, this approach can be generalized into asking what conditions affect strength of a threshold in state space. Answering this question can provide insight into mechanism.

Statistical relevance. Each empirical modeling method we test recovers data structure using a 'building material' specific to the algorithm. By analogy, CART uses square or rectangular prisms, Random Forest tends to stipple with long narrow rectangular prisms, and NPMR uses smooth, stretchy material (see Fig. 5). Model building algorithms can introduce substantial model bias when the geometric constraints of building material are not suited to the shape of the response. For example, CART's building material, square and rectangular prisms, inefficiently captures diagonal gradients. Overall, CART performs better with non-diagonal thresholds by splitting data at threshold values and creating discrete prediction levels for subsets of predictor values. This maintains square, flat areas typical of non-diagonal thresholds or thresholds responding to a single predictor (Fig. 5). CART models can be 'pruned' numerous ways (Hastie and Tibshirani 2001), which change the size of the prisms and hence sensitivity to diagonality. However, our method of pruning using ten-fold cross-validation is the most objective and robust to external data (Hastie and Tibshirani 2001), yet, this process creates large prisms.

Random Forest also uses rectangular prisms as building material but the prisms are typically much narrower and longer compared to CART. Diagonality challenges Random Forest the same way it challenges CART. Rectangular prisms inefficiently capture the diagonal faces while efficiently capturing large, rectangular, flat areas typical of non-diagonal thresholds (see Fig. 5).

NPMR produces smooth renditions of response patterns, and the sensitivity of the method to threshold strength and diagonality is not due to the geometric constraints of the predictions. The sensitivity of NPMR to threshold strength and diagonality for continuous data is likely due to the decrease in accuracy of NPMR predictions occurring when sloping surfaces abut the edges of the predictor space. The smoothing function biases the edges toward the central tendency of the data. The degree of this bias depends on the type of smoothing function and the width of the kernel per predictor. Broader kernels incur more bias.

In summary, non-parametric regression methods vary in their efficacy of capturing response shapes. They are sensitive to the threshold strength and diagonality of the underlying surface. The contribution of tests that use threshold strength and diagonality is especially relevant to the comparisons of empirical methods designed for complex data analysis such as species-habitat models in ecology (e.g. Elith et al. 2006, Guisan et al. 2007). Currently, methods are compared using real data sets of unknown structure, and the comparisons do not discern the role of the response shape in method performance (e.g. Elith et al. 2006, Guisan et al. 2007). Our work shows that non-parametric regression approaches can impose substantial model bias, and this bias depends on the geometry of the algorithm's 'building blocks' coupled with the geometry of the data structures. For example, the accuracy of CART is highest with non-diagonal shapes and lowest for diagonal shapes of data structure. The error or bias incurred from the limits of CART's algorithm is more pronounced for strongly diagonal surfaces. Strongly diagonal surfaces are not amenable to capture by rectangular prisms (the analytical type of 'building block' imposed by the algorithm). Such model bias has unknown and possibly far-reaching consequences across disciplines that apply these methods. Other disciplines using these methods range broadly from epidemiology to earth sciences.

*Methodological considerations.*We measure threshold strength and diagonality on a continuous scale rather than assigning a simple 'yes' or 'no'. Values for threshold strength can be interpreted by comparison to our benchmarks (the shapes represented among the 48 simulated data sets;Table 1, Fig. 2) or by comparison among data sets.

Because our simulated gradients represent many possible response surfaces, *T* and *D* can be applied to any ecological regression with one or two continuous predictors. *T* and *D* depend on how well the shape of the response surface is sampled and fit. All of our examples with real data involve well-sampled response surfaces with strong fits. Figure 6A and B show examples with continuous response (or dependent) variables, and Figure 6C shows a binary response variable. The indices can be used with response surfaces modeled in state space from other disciplines. An important exception includes surfaces where more than one response value corresponds to a single unique combination of input values. A classic example of this comes from the cusp catastrophe of catastrophe theory where a surface in state space exhibits a cusp-like fold in the ordinate or *z*-dimension of an *x*, *y*, *z* coordinate system (Thom 1989). Although the cusp catastrophe surface is not generated using regression, it is still a surface in state space, albeit theoretical. Folds in the ordinate dimension of state space can exist empirically and theoretically. However, regression methods cannot capture such folds, and the indices we present are not equipped to measure such folds.

Although we limit the index development to three dimensions of state space, the indices are specifically designed for ease of algebraic extension to *n*-dimensions of state space. Evaluation of multi-factor thresholds in more than three dimensions of state space would offer more realism to threshold analysis. Extension of the indices to *n*-dimensions of state space is a topic of future research. Finally, the indices are not equipped (as we present them here) to rank or measure the relative importance among predictors in eliciting a threshold in state space. However, this is can be measured using statistics from non-parametric modeling methods. For example, in NPMR, “sensitivity” is a measure of relative variable importance.

Geographic relevance. The indices of threshold strength and diagonality may conceivably be used in domains other than state space such as geographic. Thresholds in the geographic domain are considered ‘boundaries’ or transition zones that delineate patches (Cadenasso et al. 2003). Boundaries in a geographic domain can be visualized as meandering zones of abrupt change differing in extent and magnitude, and the objective

is to map and characterize these meanders across space. Employing threshold strength (presented here) incrementally within a window at a fixed resolution in geographic space may be appropriate for some applications. In fact, an algorithm measuring abruptness of geographic boundaries in ecology already exists (Bowersox and Brown 2001) based largely on the work of Fortin (1994) and Womble (1951); however, this algorithm does not provide a value of threshold abruptness that is insensitive to rotation with respect to longitude and latitude (or the analogous *x*-*y* plane). We explain this and the associated significance below.

First, boundary mapping employed by Fortin (1994) and Womble (1951) identifies abrupt change across a spatial grid of points by employing arbitrary cut-off values in the absolute values of partial first derivatives among adjacent points. The identified steep slopes and their spatial locations are called boundary elements. Bowersox and Brown (2001) build on boundary elements to develop a method to measure the abruptness of such a boundary. They measure the area under the curve representing a frequency distribution of boundary elements using a gradient of twenty different cut-off values. The idea is that strong thresholds will show a spike in numbers of boundary elements with high cut-offs. This makes a taller, narrower curve with a longer tail compared to other curves. However, partial first derivatives change across the same point pattern but rotated 45 degrees, and consequently, they are not rotationally invariant in the *x*-*y* plane. Hence, the same boundary rotated 45 degrees will yield different magnitudes of partial first derivatives tied to each boundary element. Further, the metric is not spatially explicit and does not distinguish a threshold shape from a different shape with the same frequency distribution of boundary elements.

In contrast, our threshold strength index solves these problems. Our criteria of monotonicity (Eq. 6) and bimodality (the left multiplicand Eq. 8) together describe the characteristic of the shape as a whole. The criteria distinguish abrupt thresholds from less abrupt thresholds, or abrupt thresholds from shapes with no thresholds, and so on, regardless of their orientation in the *x*-*y* plane.

CONCLUSION

Threshold strength and diagonality are measurable shape attributes of multi-dimensional thresholds. We provide new tools to quantify this underused type of information. The shape of a data pattern is fundamental to the development of theory in ecology (e.g.Whittaker 1975); yet, shortfalls in the description and understanding of a complex response shapes may be pervasive. These shortfalls can impede theoretical advancement, successful prediction, and management application (Efron and Tibshirani, 1991, Scheffer and Carpenter 2003).

We move beyond single-factor methods of quantifying thresholds that occur in state space to add realism and higher dimensionality. We introduce a parameter-free way to quantify threshold strength and diagonality from thresholds occurring in state space. Future methodological and basic research objectives for the indices include: measure if and how the prediction accuracy of other non-parametric regression methods depends on *T* and *D*, develop a roving window method that can measure the indices at different scales within a response surface, study mechanisms underlying multi-factor thresholds for ecological systems hypothesized to exhibit thresholds, and answer the question, how can this approach be used to identify systems approaching threshold responses before they happen?

ACKNOWLEDGEMENTS

We thank the Forest Inventory and Analysis program of the U.S.D.A. Forest Service for funding this work. We thank the following people for data or comments: Peter Dolan, Paul Murtaugh, Alix Gitelman, and Richard Waring.

FIGURES

Figure 1. (**A**) We calculate three-dimensional threshold strength from a modeled surface formed collectively by a grid of points. A 'spider' is established for each unique set of nine adjacent points indexed as shown. The circular spider on the right results from interpolating the diagonal vectors in the square spider on the left. Each *z* represents a response point, and each *z** represents an interpolated response point. Four pairs of opposing vectors are defined for each spider, *NESW*, *NWSE*, *NS*, and *EW*. (**B**) We demonstrate calculation of diagonality from a modeled three-dimensional surface formed collectively by many four-sided polygons defined by points as shown; the diagram to the upper right represents one polygon and illustrates indexing of the four points for calculating a metric, *d*, for each polygon to sum across the surface.

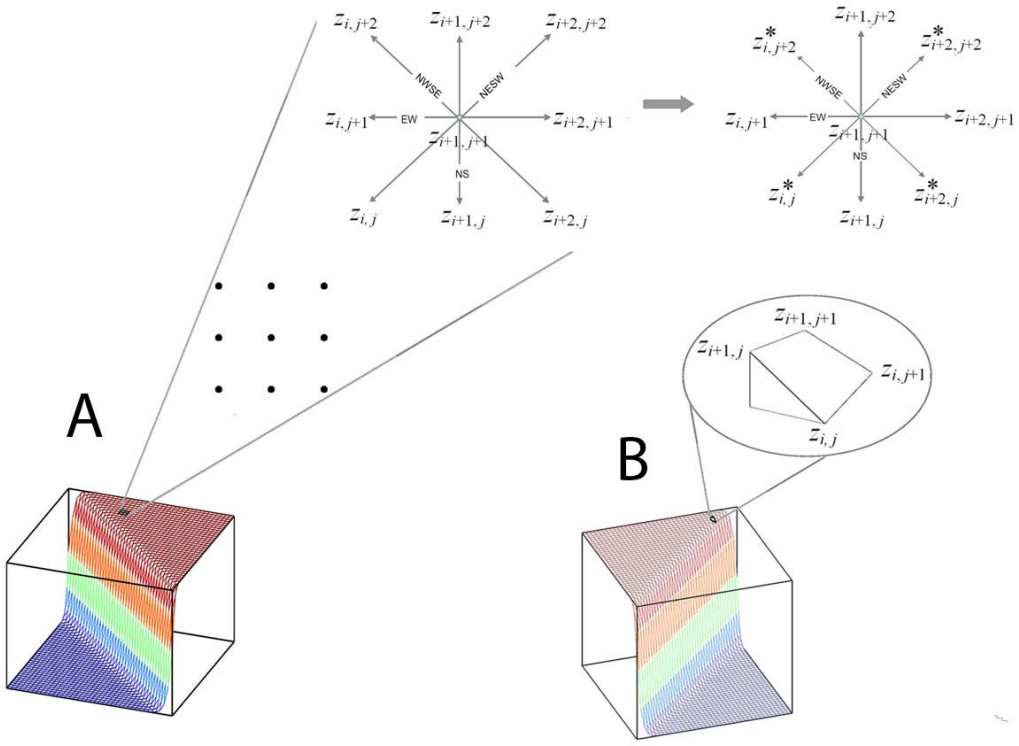


Figure 2. Bird's eye views of three-dimensional 48 simulated response surfaces. Each surface is labeled to match corresponding names and index values in Table 1. The color gradient represents different values of for each response ranging from min to max as shown.

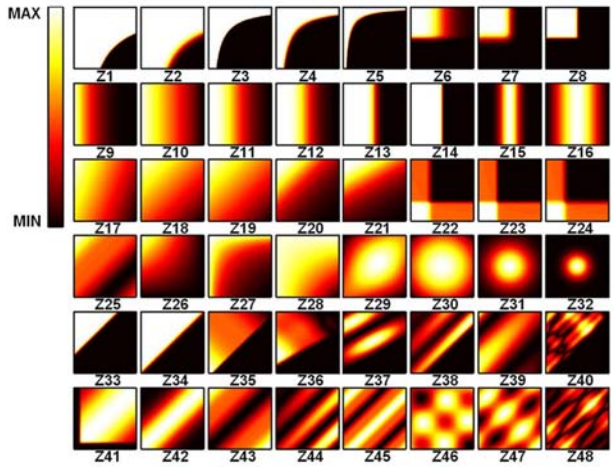


Figure 1 Legend

Label	Pattern name	Label	Pattern name
Z1	steep Niagara step off-center	Z25	short stream valley
Z2	Niagara step off-center	Z26	$z=xy$
Z3	steep Niagara step	Z27	Niagara kite
Z4	Niagara step	Z28	$z=x-xy$
Z5	Niagara edge step	Z29	diag Gaussian hill-ridge
Z6	corner wave step	Z30	parachute
Z7	corner square step	Z31	Gaussian wide hill
Z8	steep corner square step	Z32	Gaussian hill
Z9	$z=x^3$	Z33	steep diag step off-center
Z10	$z=x$	Z34	steep diag step
Z11	s-curve	Z35	skate-ramp abrupt diag
Z12	gentle step	Z36	Gaussian skate-ramp abrupt
Z13	moderately steep step	Z37	diag two waves abrupt
Z14	steep step	Z38	zig-zag diag abrupt
Z15	Gaussian ridge	Z39	wide zig
Z16	wide Gaussian ridge	Z40	organic peaks
Z17	$z=x+0.4y$	Z41	diag Gaussian ridge abrupt L
Z18	$z=x+y$	Z42	diag Gaussian ridge
Z19	$z=0.5x+0.5y$	Z43	stream valley
Z20	s-curve off-center 45	Z44	zig-zag slant
Z21	s-curve off-center 60	Z45	diag waves
Z22	triple staircase diff levels	Z46	four eggs
Z23	steep triple staircase	Z47	undulating sea
Z24	triple staircase	Z48	weaving

Figure 3. (**A**) Four three-dimensional planes (top row) with each generating function (titled above) and matching bird's eye views (bottom row). The planes increase in diagonality from left to right. Perfectly diagonal surface area (e.g. the right-most plane) represents equivalence among partial first derivatives. In contrast, non-diagonal surface area varies strictly with one predictor (e.g. the left-most plane). In between the extremes, the rate of change of *z* with respect to *y* gradually becomes more important until it reaches equivalence with the rate of change of *z* with respect to *x* for a diagonal plane. (**B**) Four three-dimensional response structures (top row) with each generating function (titled above) and matching bird's eye views (bottom row). Each function represents a different additive, statistical model containing an interaction term or, in this case, a multiplicative term comprising predictors *x* and *y*. The details of the four additive models vary; in particular, the functions are a mix of different orders. The four structures capture the shape family of additive models with multiplicative terms, hyperbolic paraboloids. Despite the details of the additive model and multiplicative terms, the equations yield similar shapes. Interactions of this sort create some diagonality but strongly diagonal surfaces such as diagonal planes need not have statistical interactions.

Figure 3 Continued.

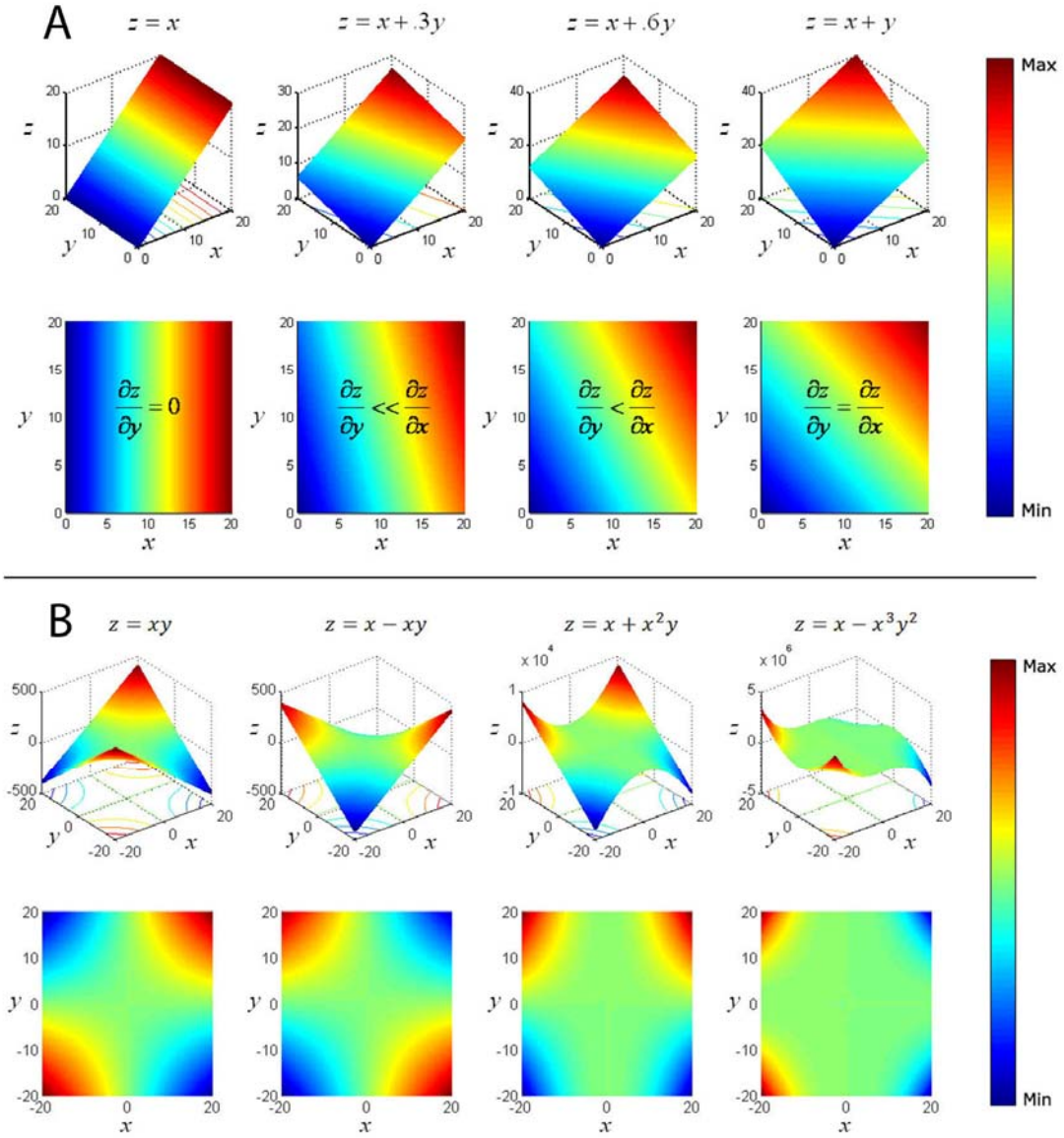


Figure 4. Prediction accuracy of Random Forest, Classification and Regression Trees (CART), and Non-parametric multiplicative regression (NPMR) for continuous and binary responses plotted versus threshold strength (*T*) and diagonality (*D*). Each point is a median, externally-validated accuracy of 100 models built from random samples size *N*=250. The samples draw from a large subset of a simulated data set (*N*=9,201). A separate subset of 1000 points performs external validation for each simulated data set (see Fig. 2 for the simulated data sets). Error bars represent 95% quantiles.

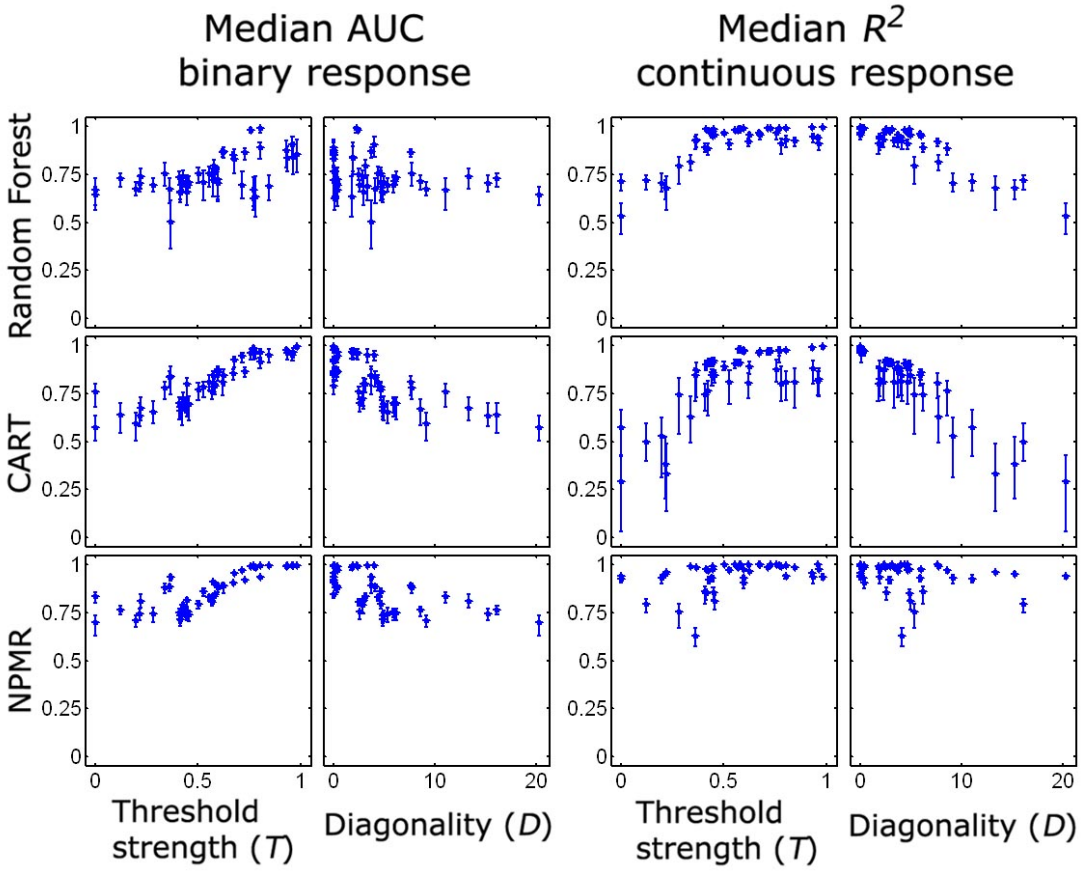


Figure 5. A visual comparison of predicted three-dimensional response surfaces together with surfaces showing the original data (bird's eye views). Results are shown from three modeling methods for two types of responses, continuous and binary. In the case of the right-most three columns, the binary response surfaces represent the probabilities of underlying point densities. The top row depicts the original response surfaces each comprising 10,201 data points or 100 by 100 increments. The lower rows show predicted values for models built from a random subsample (size *N*=250) of each original response surface in the top row. Two replications of random samples are shown for each modeling method to provide a sense of the variation.

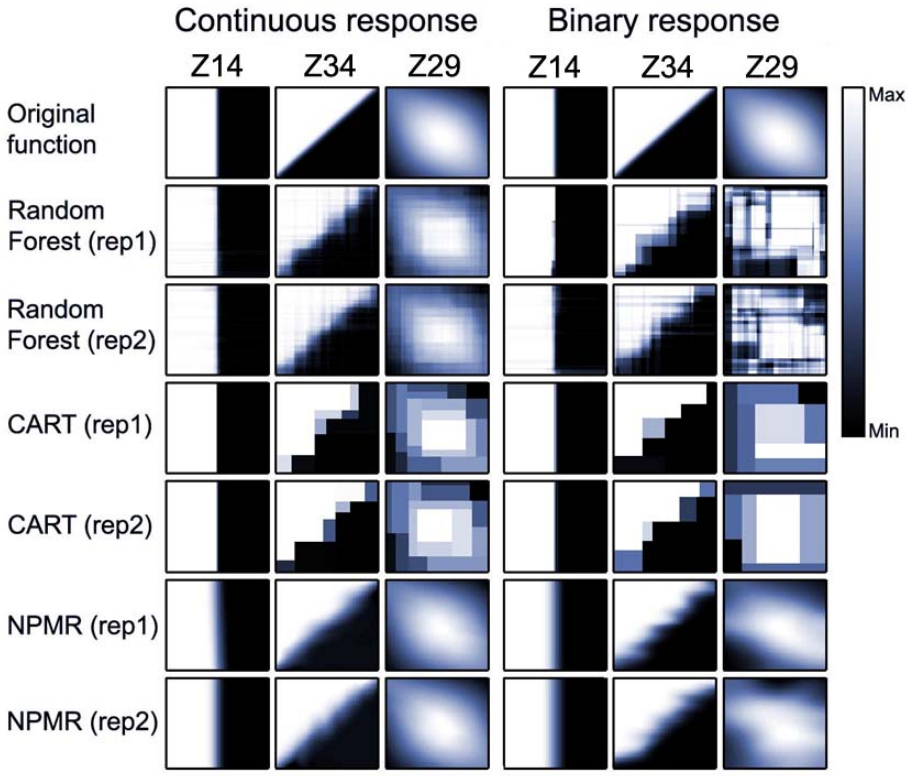
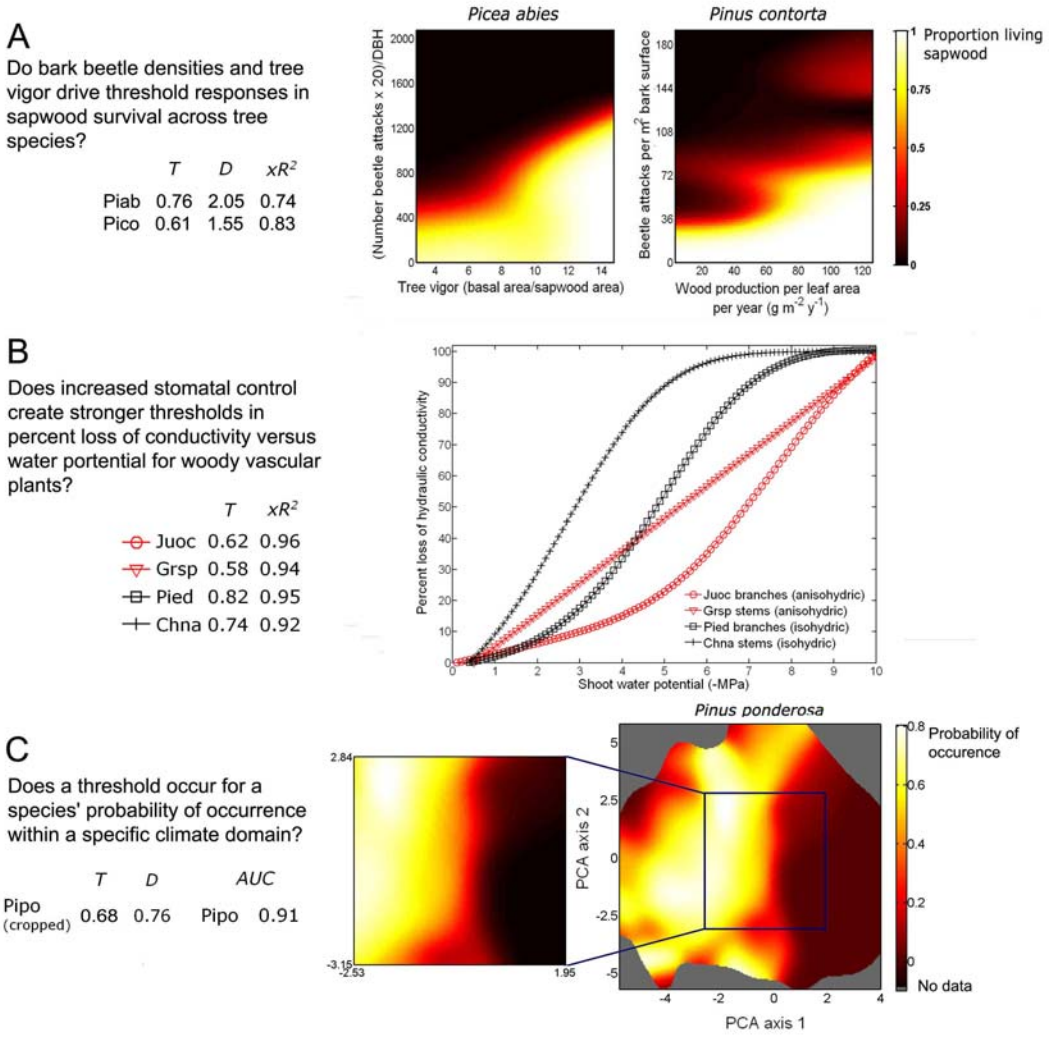


Figure 6. The application of threshold strength (*T*) and diagonality (*D*) to ecological data using prediction surfaces (right) generated by NPMR for different data sets (see Appendix C: Table C1 for specifications). Threshold strength (*T*) and diagonality (*D*) values (middle) are measured from modeled surfaces (right) test different questions (left). (**A**) Proportion of sapwood survival versus separate measures of tree vigor and severity of bark beetle attack for two tree species. Left: *Picea abies* attacked by *Ips typographus.* Right: *Pinus contorta* attacked by *Dendroctonus ponderosae*; data and theory from (Christiansen et al. 1987). (**B**) Mean percent loss in hydraulic conductivity versus shoot water potential (-MPa) of branches of two isohydric (red lines) and two anisohydric (black lines) species. Lines show predicted curves from data of stems of two shrub species, *Grayia spinosa* and *Chrysothamnus nauseosus* (Hacke et al., 2001), and branches of two tree species *Juniperus osteosperma* and *Pinus edulis* (Linton et al. 1998) (see Appendix B: Fig. B1). (**C**) Probability of occurrence for a dominant tree species in Oregon, *Pinus ponderosa*, modeled relative to two axes derived using Principal Components Analysis representing a summer aridity gradient (Axis 2) increasing vertically and a continentality gradient (Axis 1) increasing from right to left (see Appendix B: Fig. B2). Together, the axes explain 82% of variability in the source data (see Appendix C: Table C2 for table of source data). Threshold strength and diagonality are measured from a cropped portion of the surface (left). Data come from 1724 plots of the Forest Inventory Analysis program in Oregon (Azuma et al. 2002, 2004).

Figure 6 Continued.



TABLES

Table 1. Simulated data surfaces from Figure 2 named and ranked in descending order by values of diagonality (*D*) (left panel) and threshold strength (*T*) (right panel).

D	Label	Surface name	*T*	Label	Surface name
20.31	Z48	weaving	0.980	Z14	steep step
16.15	Z45	diag waves	0.963	Z3	steep Niagara step
15.29	Z44	zig-zag slant	0.961	Z34	steep diag step
13.36	Z38	zig-zag diag abrupt	0.936	Z4	Niagara step
11.03	Z40	organic peaks	0.931	Z13	moderately steep step
9.12	Z47	undulating sea	0.846	Z33	steep diag step off-center
8.58	Z43	stream valley	0.805	Z1	steep Niagara step off-center
7.71	Z37	diag two waves abrupt	0.801	Z12	gentle step
7.63	Z42	diag Gaussian ridge	0.781	Z5	Niagara edge step
6.17	Z39	wide zig	0.772	Z8	steep corner square step
5.96	Z30	parachute	0.770	Z7	corner square step
5.90	Z25	short stream valley	0.757	Z2	Niagara step off-center
5.31	Z46	four eggs	0.727	Z11	s-curve
4.93	Z29	diag Gaussian hill-ridge	0.716	Z6	corner wave step
4.85	Z18	$z=x+y$	0.676	Z15	Gaussian ridge
4.85	Z19	$z=0.5x+0.5y$	0.673	Z16	wide Gaussian ridge
4.65	Z35	skate-ramp abrupt diag	0.623	Z41	diag Gaussian ridge abrupt L
4.64	Z31	Gaussian wide hill	0.618	Z42	diag Gaussian ridge
4.42	Z20	s-curve off-center 45	0.597	Z24	triple staircase
4.14	Z36	Gaussian skate-ramp abrupt	0.597	Z22	triple staircase diff levels
4.05	Z34	steep diag step	0.596	Z23	steep triple staircase
3.73	Z32	Gaussian hill	0.583	Z10	$z=x$
3.69	Z41	diag Gaussian ridge abrupt L	0.576	Z21	s-curve off-center 60
3.35	Z33	steep diag step off-center	0.575	Z9	$z=x^3$
3.14	Z21	s-curve off-center 60	0.569	Z27	Niagara kite
2.99	Z26	$z=xy$	0.554	Z20	s-curve off-center 45
2.9	Z28	$z=x-xy$	0.528	Z35	skate-ramp abrupt diag
2.55	Z17	$z=x+0.4y$	0.500	Z31	Gaussian wide hill
2.52	Z27	Niagara kite	0.454	Z29	diag Gaussian hill-ridge
2.37	Z2	Niagara step off-center	0.449	Z17	$z=x+0.4y$
2.25	Z1	steep Niagara step off-center	0.448	Z30	parachute
1.94	Z4	Niagara step	0.445	Z26	$z=xy$
1.89	Z3	steep Niagara step	0.443	Z25	short stream valley
1.77	Z5	Niagara edge step	0.426	Z28	$z=x-xy$
0.37	Z6	corner wave step	0.421	Z43	stream valley
0.36	Z24	triple staircase	0.412	Z18	$z=x+y$

Table 1 continued.

D	Label	Surface name	*T*	Label	Surface name
0.34	Z7	corner square step	0.412	Z19	$z=0.5x+0.5y$
0.26	Z22	triple staircase diff levels	0.409	Z39	wide zig
0.21	Z23	steep triple staircase	0.365	Z32	Gaussian hill
0.10	Z8	steep corner square step	0.362	Z36	Gaussian skate-ramp abrupt
0.00	Z14	steep step	0.336	Z37	diag two waves abrupt
0.00	Z12	gentle step	0.280	Z46	four eggs
0.00	Z13	moderately steep step	0.220	Z38	zig-zag diag abrupt
0.00	Z11	s-curve	0.215	Z44	zig-zag slant
0.00	Z16	Gaussian ridge	0.198	Z47	undulating sea
0.00	Z15	wide Gaussian ridge	0.122	Z45	diag waves
0.00	Z10	$z=x$	0.000	Z40	organic peaks
0.00	Z9	$z=x^3$	0.000	Z48	weaving

APPENDIX A: SUPPORTING METHODS

Threshold strength calculation in two dimensions. Here, we present equations to calculate threshold strength from two-dimensional, modeled surfaces (e.g. Fig. 6B). The response values are divided by their maximum range to standardize among response surfaces with different ranges as in three dimensions. Spiders comprise $3^{(w-1)}$ points where *w* is number of dimensions. This yields a spider of three points in two dimensions (for a single predictor divvied into 100 increments or 101 points). A single metric measuring departure from monotonicity, *EW*, is subject to the following criterion as in three dimensions: if the two endpoints are either both above or both below the center point, then departure from monotonicity occurs, if not, then departure from monotonicity is zero. For cases expressing departure from monotonicity, the degree of the departure follows

$$EW_i = \min\left\{\left|z_i - z_{i+1}\right|, \left|z_{i+2} - z_{i+1}\right|\right\}. \quad \text{(A.1)}$$

EW represents departure from monotonicity for the spider in two dimensions

$$S_i = EW_i. \quad \text{(A.2)}$$

The sum of all *S* or $\sum_{i=1}^{N-2} S_i$ becomes the overall departure from monotonicity for the response surface. The average departure from monotonicity (*K*) derives from

$$K = \frac{\sum_{i=1}^{N-2} S_i}{N-2} \quad \text{(A.3)}$$

where the denominator is the total number of vector contrasts evaluated for the response surface, *N*-2, or 99 for the fixed scale of 100 increments or 101 response points. We calculate monotonicity (*M*) using a negative exponential function, e^{-950K}, as in three dimensions. Threshold strength (*T*) is the product of monotonicity (*M*) and the bimodality of the response (the left multiplicand in Eq. A.5) as in three dimensions

$$\sigma_z = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(z_i - \bar{z})^2}, \quad \text{(A.4)}$$

$$T = 2\sigma_z M \ . \tag{A.5}$$

Linear interpolation of NWSE and NESW for the spider in three dimensions. The 3D spider yields four pairs of opposing vectors with each vector spanning two grid points, the outer point and the center (Fig. 1). Diagonal vectors are shortened (z^*) to match the radius of non-diagonal vectors through linear interpolation following

$$z^* = z_{outer} + (\sqrt{2} - 1)(z_{center} - z_{outer}) \ . \tag{A.6}$$

Equation A.6 calculates a single interpolated point (Fig. 1). The center and outer points and are illustrated in the diagonal 'legs' of the spider (Fig. 1). The resulting points yield a spider with a circular footprint. We discovered that this method of interpolation (using two points) yields values essentially equivalent to other methods incorporating more than two points because we apply it to smooth surfaces at a fixed, fine resolution.

Testing threshold strength across dimensions. Consider response surfaces derived using a four-parameter form of the logistic function, which generates surfaces resembling steps but of variable steepness

$$z = \frac{a - d}{1 + \left(\frac{x}{c}\right)^b} \tag{A.7}$$

where *a* is the maximum horizontal asymptote, *d* is the minimum horizontal asymptote, *c* is the value of *x* at the inflection point, and *b* controls the steepness of the step. If we vary *b* to increase the steepness of the step with all else equal, the threshold strength (*T*) of a surface should increase. We calculate threshold strength for two-dimensional and three-dimensional threshold surfaces using Equation A.7 while varying only dimensionality of the input. We assume a maximum horizontal asymptote (*a*) of 30, minimum asymptote (*d*) of 0, and inflection point (*c*) of 10. The input for the three-dimensional surfaces is a perfect grid of 101 by 101 points where *x* and *y* each range from 0 to 20 (see Appendix B: Fig. B3). Two-dimensional curves follow 100 increments of *x* from 0 to 20. Threshold strength (*T*) increases with *b* or steepness in the rise of the step as shown in Figure B3 of Appendix B. Threshold strength values for the three-dimensional and two-dimensional

responses generated from same underlying logistic equation are equivalent (see Appendix B: Fig. B3).

Testing diagonality by rotation. Consider a single surface rotated in the *x-y* plane. Diagonality (*D*) should decrease with the angle of the response gradient relative to one reference predictor axis (e.g. *x*) along 45 degrees of rotation. To test this, we translate the surface origin to the center of the predictor space (10, 10). The function input ranges from 0 to 20 per predictor on a grid of 101 by 101 increments. We trim the base to a circle to facilitate uniformity in sample size and shape each time the circle is rotated (see Appendix B: Fig. B4); rotating a square base within the constraints of a grid can yield uneven edges or slight sample size differences among rotations. Diagonality decreases with the angle of the response gradient relative to one reference predictor axis (e.g. *x*) along 45 degrees of rotation (see Appendix B: Fig. B4).

Converting continuous response to binary with simulated data. We simulate 48 binary response surfaces with the same underlying structure as the 48 continuous response surfaces. To do this, we treat the continuous response as a probability structure that converts to a point density for a binary response. First, we create a vector (or string of numbers), **c'**, for each continuous response vector, **c**, composed of random numbers drawn from a uniform distribution within the same range as **c**. Values for elements of **c'** follow $c_i^{'} = \min(c_i) + (\max(c_i) - \min(c_i)) * rnd$ where *rnd* is a uniformly distributed random number ranging from 0 to 1, and c_i is an element of the continuous response vector, **c**, and *i* indexes the vector position. We execute a conditional rule among vectors to create the binary vector: if $c_i > c_i'$, then b_i =1; otherwise, if $c_i \leq c_i'$, then b_i =0 with **b** as the resulting binary vector.

Explanation of the Area Under the Receiver Operator Characteristics (ROC) curve (AUC). The AUC is a metric derived from an ROC curve for the purpose of comparing the performance of probabilistic classifiers. Probabilistic classifiers are algorithms yielding instance probabilities or the degree to which an instance belongs to a class rather than output in the form of full class membership. The ROC curve is designed to visualize such a classifier's performance by plotting the *true positive rate* versus the

false positive rate for different cut-off levels. A cut-off level converts the instance probability to a category. For example, in the case of a binary response, a cut-off level of 0.7 would classify an instance probability of 0.8 as a 1 and an instance probability of 0.5 as a 0. The *true positive rate* at each cut-off level is calculated as the total number of positives correctly classified divided by the total of the positives classified. The *false positive rate* at each cut-off level is calculated as the total number of negatives incorrectly classified divided by the total of the negatives classified. Points corresponding to a gradient of cut-off values are plotted to form the ROC curve. The area under this curve is calculated to give a measure of performance that does not depend on either a single cut-off value or the proportion of class representation (the frequency of 1s compared to 0s in the original data) (Fawcett 2006). The 1:1 line on the graph represents an AUC of 0.5 or performance no better than random chance.

Explanation and implementation of CART. CART creates a dichotomous tree model with a series of splitting rules using binary recursive partitioning. Each branch split in the tree results from minimizing the variability of two separate groups of data. A predictor at a single value serves as a data partition or branch split. Data are successively partitioned until branch endpoints are homogenous or a default number of data points in branch endpoints are reached, whichever comes first. The number of terminal nodes or endpoints of the tree gauge the tree size, and each terminal node yields an average prediction for regression or the most probable category for classification (categorical response). A CART tree requires 'pruning' or cutting branches from the tree (Hastie and Tibshirani 2001). The most objective and robust method of pruning is achieved by cross-validation, and we performed automated ten-fold cross-validation to guide pruning for each CART model we present (Hastie and Tibshirani 2001). Cross-validation is a process where a random division of data into two groups creates a 'testing' set and 'training' or 'learning' set. A CART model grows from the training set and prediction accuracy for the model is measured with the testing set. Data division and model validation occur repeatedly at numerous tree sizes, and the lowest average variability (or deviance in this case) per tree size determines the optimal tree size. We implement the CART analysis for

continuous and binary data surfaces using the computer language S in the software package S-Plus 7.0. Chambers and Hastie (1992) described the code to implement CART and perform the cross-validation.

Explanation and implementation of Random Forest. The Random Forest analysis generates many CART-like trees where each tree grows from a random sub-sample of data rather than the whole pool. Additionally, a random selection of predictors determines the possible candidates to define each split of each tree, and no pruning is performed on any trees. Test values are run through the entire forest and a predicted value is either a vote count from all trees (for classification) or an average value from all trees (for regression). Berk (2006) provides excellent introductions to Random Forest and CART.

We implement Random Forests in computer language R, which is available in the software package R 2.3.1. We use default settings of the function "randomForest," which builds the forest from 500 trees. The function and its default settings are described in help text of the "randomForest" package available at http://cran.cnr.Berkeley.edu/bin/windows/contrib/2.3/randomForest_4.5-16.zip.

Explanation and implementation of Non-Parametric Multiplicative Regression (NPMR). NPMR is a kernel smoothing technique that objectively determines kernel width by maximizing fit from cross validation. It works like other kernel smoothers where local windows center on target points within the predictor space. A weight function (or kernel function) applies in a window to relegate importance or numeric weight to points closest to the target. A local model such as weighted linear regression predicts the dependent variable at the target point. A prediction curve or surface of estimates results from repeating this process for all target points. NPMR differs from other kernel smoothing applications in several aspects. First, NPMR optimizes the breadth of the smoothing function for each predictor, known as the kernel width or tolerance, to maximize global fit. Second, NPMR leaves the target point out when predicting the target within a window to ensure robust prediction, which is known as leave-one-out cross validation. Third, NPMR multiplies weights from individual predictors to accommodate complex interactions in many dimensions. NPMR provides

measures of model fit for model selection that incorporates the leave-one-out cross-validation, namely cross-validated R^2 for a continuous response and AUC and log likelihood ratio (Log*B*) for a binary response. We automate model fitting for random sampling using NPMR default settings.

APPENDIX B: SUPPORTING FIGURES

Figure B1. Scatter plots and predicted curves of the mean percent loss in conductivity versus shoot water potential by species for data presented in (Linton et al. 1998, Hacke et al. 2000) and modeled using NPMR (Fig. 6B). NPMR specifications for Figure 6 are listed in Table C1 of C.

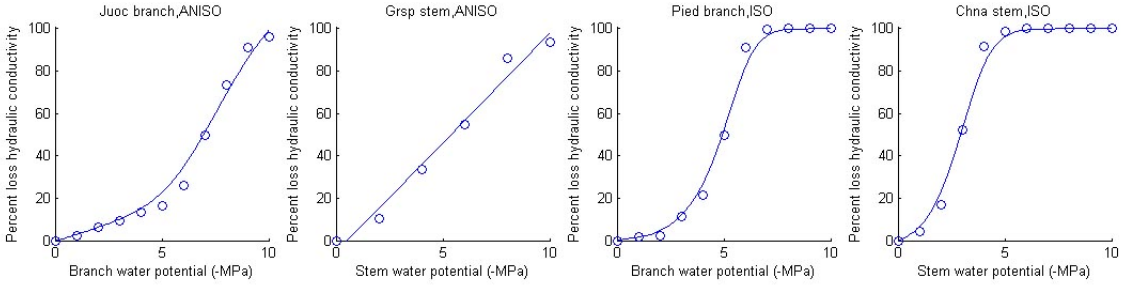


Figure B2. Ordination of first two axes from a Principal Components Analysis of variables defined in Table C2 of Appendix C. Axis 1 and 2 serve as predictors for models of tree species' presence in Figure 6C. Each point represents a plot from the Forest Inventory Analysis of the U.S. Forest Service; 1724 plots sample a systematic grid of density 3.4 miles in eastern and western Oregon (Azuma et al. 2002, 2004). The distance from one point to the next in the ordination is proportional to their dissimilarity with respect to the variables shown. The angle and length of lines in the ordination represent loadings of variables on the PCA axes. Axis 1 captured a continentality gradient while axis 2 captured a summer drought gradient. The axes explain 82% of the variability in the data.

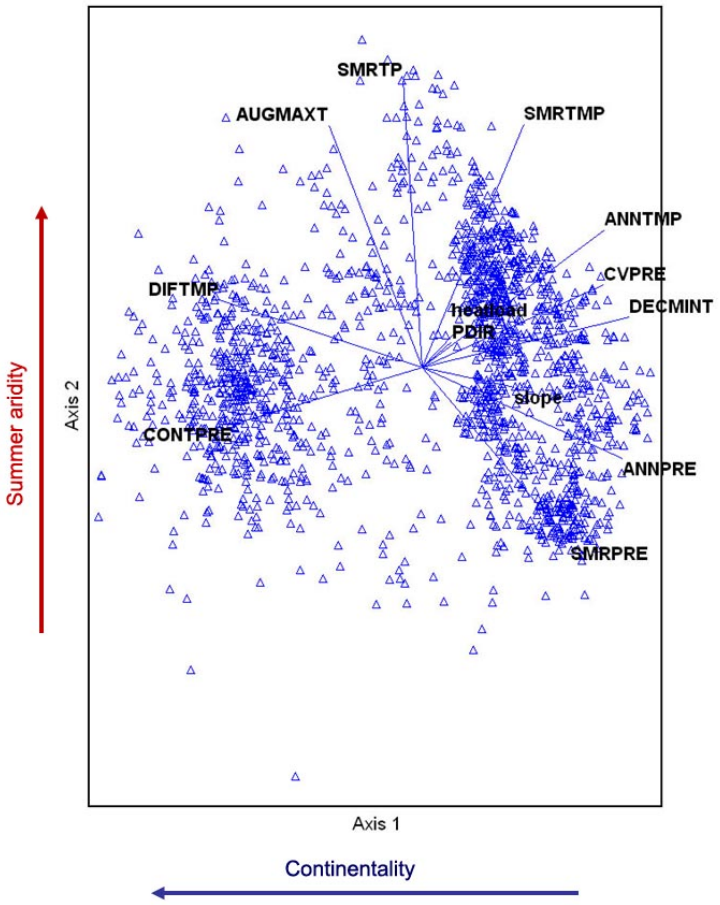


Figure B3. Threshold strength (*T*) plotted versus the logistic function exponent, *b* (determining steepness), for three-dimensional patterns as black circles and two-dimensional patterns as red crossmarks (Panel A). The same equation with differing dimensionality of input was used to generate the three-dimensional versus two-dimensional patterns. Panel B shows corresponding three-dimensional patterns for select values of *b* plotted in Panel A.

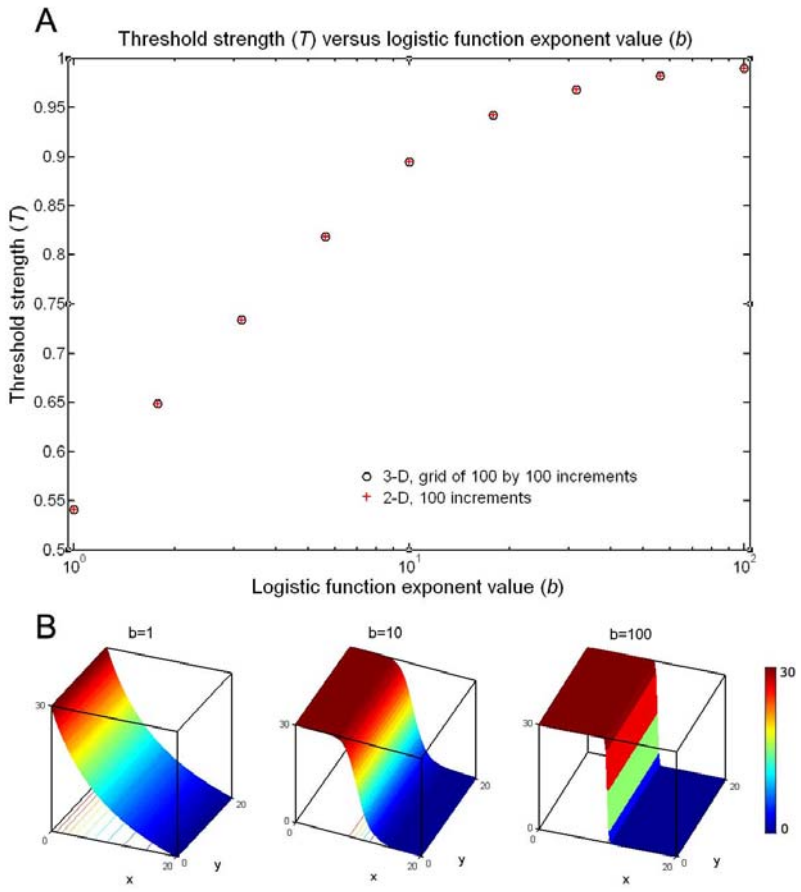
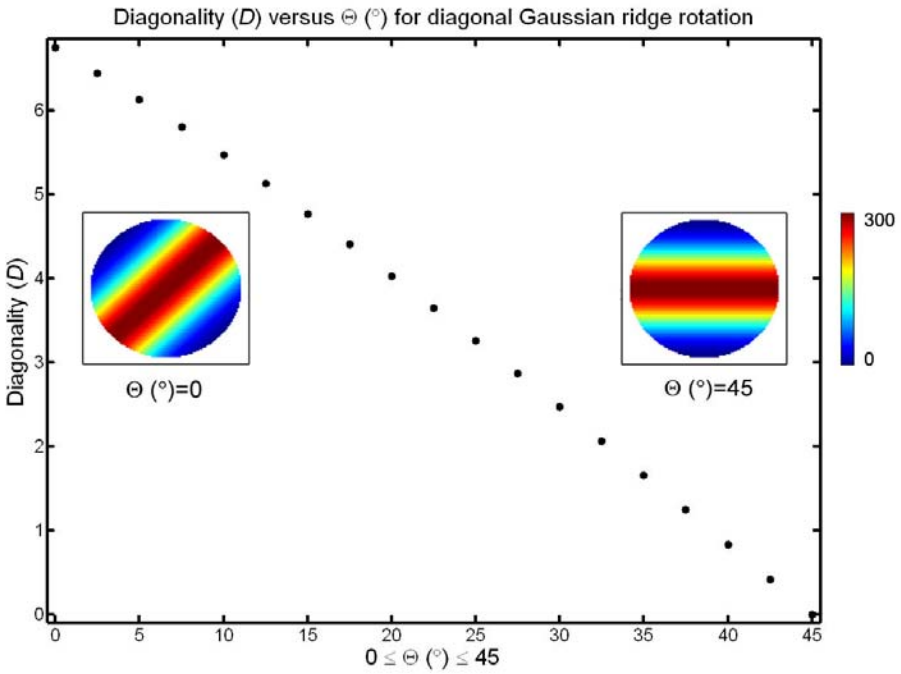


Figure B4. Diagonality (*D*) decreases as the three-dimensional diagonal Gaussian ridge (shown) is rotated θ degrees in the (*x*,*y*) plane. The three-dimensional surfaces depict the diagonal Gaussian ridge for each extreme value of θ. The surfaces sit on a circular base.



APPENDIX C: SUPPORTING TABLES

Table C1. NPMR specifications and results for models displayed in Figure 6. NPMR is further described by McCune (2006). Specifications are associated with settings from Hyperniche version 1.19 (McCune and Mefford 2004). Table abbreviations are as follows: LM is local mean model form, LL is local linear model form, Q represents a quantitative (or continuous) response, B is a binary response, *N** is the average neighborhood size, tolerance is the kernel width per predictor, and PLC is the percent loss of hydraulic conductivity.

Table C1 Continued.

Response	Model form	Data type	*N**	Predictor	Tol-erance
Panel A					
Picea abies; sapwood survival	LM	Q	10.8	Wood production	2.2
				(Bark beetle attacks*20)/DBH	207.3
Pinus contorta; sapwood survival	LM	Q	7.6	Tree vigor	22.2
				Bark beetle attacks/m^2	11.5
Panel B					
Juniperus osteosperma; PLC	LL	Q	2.4	Branch water potential (-MPa)	1.5
Grayia spinosa; PLC	LL	Q	1.6	Stem water potential (-MPa)	8
Pinus edulis; PLC	LL	Q	0.9	Branch water potential (-MPa)	0.8
Chrysothamnus nauseosus; PLC	LL	Q	0.9	Stem water potential (-MPa)	0.8
Panel C					
Pseudotsuga menziesii presence/absence	LM	B	235	PCA axis 1	0.88
				PCA axis 2	1.18
Pinus ponderosa presence/absence	LM	B	98	PCA axis 1	0.49
				PCA axis 2	0.66

Table C2. Names and definitions of variables used to derive principal components for Figure 6C. Most variables and definitions follow use by Ohmann and Spies (1998). We use Daymet climate data (Thornton et al. 1997), and latitude and slope data for radiation and heatload estimates following McCune and Keon (2002). See Appendix B, Fig. B2, for the accompanying PCA ordination.

Variable	Definition
ANNPRE	Natural logarithm of mean annual precipitation (mm)
SMRPRE	Natural logarithm of mean precipitation from May to September (mm)
CVPRE	Coefficient of variation of mean monthly precipitation during wet and dry months (December and July)
SMRTMP	Mean summer temperature (^{o}C) from May to September
SMRTP	Moisture stress during the growing season; a ratio of mean summer temperature (SMRTMP) over mean summer precipitation (SMRPRE)
ANNTMP	Mean annual temperature (^{o}C)
AUGMAXT	Mean maximum temperature in August (^{o}C)
DECMINT	Mean minimum temperature in December (^{o}C)
DIFTMP	Difference between AUGMAXT and DECMINT (^{o}C)
CONTPRE	Percentage of mean annual precipitation falling June through August
slope	Average slope of the plot (%)
PDIR	Potential annual direct incident radiation (MJ cm^{-2} yr^{-1})
heatload	Index accounting for aspect with zero being coolest and one as warmest (S6).

Chapter 3. Sensitivity of Climatic Niche Models to Changes in Inventory Method

Heather E. Lintz, Andrew N. Gray, Bruce McCune

ABSTRACT

Data from large-scale biological inventories are essential for understanding and managing Earth's ecosystems. The Forest Inventory and Analysis Program (FIA) of the U.S. Forest Service is the largest biological inventory in North America; however, the FIA inventory recently changed from an amalgam of different approaches to a nationally-standardized approach in 2000. The two types of inventories represent different probabilities of detecting trees per sample unit. Published analyses that rely on FIA data may lump FIA data from different regionally-based designs (pre-2000) or lump data across the temporal changeover without exploring the consequences. The main goal of this study is to evaluate the effect of the inventory approach on a common analysis in ecology, modeling of climatic niches (or species-climate relations). We use non-parametric multiplicative regression (NPMR) to build and compare niche models for 41 tree species from the old and new FIA design in the Pacific coastal United States. We discover two likely effects of differences in inventory approach on niche models and their predictions. First, there is an increase from 4 to 6% in random error; this is noted for modeled predictions from the different inventories when compared to modeled predictions from two samples of the same inventory. Second, systematic error (or directional disagreement among modeled predictions) is detectable for 4 out of 41 species among the different inventories: *Calocedrus decurrens*, *Pseudotsuga menziesii*, and *Pinus ponderosa*, and *Abies concolor*. Hence, at least 90% of niche models and predictions of probability of occurrence demonstrate no obvious effect from the change in inventory design. Further, the fit or accuracy of all models developed for species' occurrence based on climate was high for both data sets (mean externally-validated AUC = 0.935). Tree species' occurrence is generally and strongly linked to climate in the Pacific coastal United States.

INTRODUCTION

Survey data collected *in situ* across space and time are indispensable for understanding and managing earth's ecosystems. Biological inventories occur worldwide as repositories of ecological data that can accommodate diverse stakeholders and research (EC 1997, Rudis 2003a,b). The Forest Inventory and Analysis Program (FIA) of the U.S. Forest Service conducts the largest *in situ* forest data collection effort in North America. However, the current nationally-standard inventory resulted from recent modifications of regional sample designs beginning in 2000. Before 2000, forest inventory data measured by FIA used varied plot sizes, densities, sampling extents, sampling periods, and protocols. While these differences do not affect the ability to provide statistical summaries at local to national scales for many attributes of interest (Barrett 2004, Bechtold and Patterson 2005), the comparison of plot-level attributes across inventories can be affected (Gray 2003). To use historical data with new standardized data, we need to confront an important issue that can pervade other biological inventories and large scale data (e.g. Nelson et al. 1990, Hijmans et al. 2000, NRC 2000): how do differences among inventories, when combined, confound patterns found in the data. While one study concluded that lumping FIA sample designs is hazardous for the assessment of tree migration through time (Woodall et al. 2009a), data from the different FIA sample designs are often lumped across regions (before 2000) or across sample designs (pre- and post-2000) without knowledge of the consequences (e.g. Ohmann and Gregory 2002, Smith et al. 2004, Swenson and Waring 2006).

One important use of forest inventory data is the development of ecological niche models (e.g. Iverson and Prasad 1998, Rehfeldt et al. 2008, McKenzie et al. 2003, Svenning and Skov 2004, Evans and Cushman 2009). The relationship between a species and its environment is part of a 'species niche' or an '*n*-dimensional hypervolume' that describes conditions where a species can persist (Hutchinson 1957). Currently, niche models are used for many research purposes including species' conservation (e.g. Hannah et al. 2002, Marini et al. 2009), species' re-introduction (Yanez and Floater 2000), species' migration and invasion (e.g. Woodall et al. 2009b, Crossman et al. 2011),

biodiversity conservation (e.g. Newbold and Eadie 2004), specimen collection (e.g. Jarvis et al. 2005), and the discovery of new species (Raxworthy et al. 2003). Niche models are also used to investigate basic scientific questions on various topics (e.g. Graham et al. 2004, Kelly et al. 2008, Engelbrecht et al. 2007, Svenning and Skov 2004, Hugall et al. 2002).

Generally, the effect of combining data collected using different approaches is a problem not only for niche modeling but for other applications, data sets, disciplines, and regions. For FIA data, differences in sample effort result in different plot-level probabilities of detecting species (Grosenbaugh and Stover 1957). This can bias niche models, particularly if different sample designs are used in different portions of a species' niche and/or geographic range. In addition, sample effort typically increases with tree size in forest inventories, which could bias models if tree size varies substantially in different portions of a species' niche or range. Consequently, we ask, does the change in FIA sample design affect niche models and their predictions for tree species across the Pacific coastal United States? In so doing, we determine how results from an amalgam of approaches (from the old inventory) compare to results from a single, large-scale standardized approach (the new inventory). We assume little change in species' probability of occurrence across two sequential time periods of study. We also treat all aspects of sampling that changed in 2000 (further described below) as a single source of potential error. We compare data from the old and new design. We also compare two samples of data from the new design. We juxtapose these comparisons to identify error due to sub-sampling within a given data set. Specifically, we ask:

-How does the change in FIA inventory affect random error among predictions for models built from each inventory?

-How does the change in FIA inventory affect systematic error among predictions for models built from each inventory?

-How do random and systematic errors in modeled predictions compare among samples *within* the new, standardized inventory and *among* samples across inventories?

-How do models and geographic maps of modeled predictions compare among samples across inventories and within the new, standardized inventory?

We answer these questions and also investigate evidence of bias in the data themselves to assist our interpretation of its effect on models and predictions.

METHODS

FIA inventories. The U.S. Forest Service Research branch was mandated in 1928 to report on the status and trends of forest resources on all lands, with a focus on timber production (USDA 1992). Since the 1930s, surveys have differed by state and region, and inventories were conducted all-at-once for a state (known as the periodic inventory). The periodicity of state inventories ranged from 4 to 18 or more years, and by the early 1990s, most states completed a third inventory cycle with some states re-inventoried as many as six times (Hiserote and Waddell 2003, USDA 1992). Starting in 1990, the Forest Service initiated surveys through the Forest Health Monitoring Program (FHM), which used a fixed sampling grid across the U.S. with a quarter of the grid measured each year (Scott et al. 1993). A decade later, the periodic inventory merged with FHM to form the nationally-consistent annual inventory. Instead of periodic sampling that differed by state and rotated by state, the annual inventory samples 10% of permanent plots yearly in the west on a common national grid irrespective of state boundaries.

Key differences among the periodic and the annual inventories for our study region include periodicity, grid density, sampling extent, plot size, and protocol (Tables 1 and 2). Each inventory represents a systematic sample with variable density, sampling dates, and types of lands included. In the old periodic inventory, some management regions only included land capable of timber production (or sites capable of producing 1.4 cubic meters per hectare per year at their peak of mean annual increment). Not all inventories included lands protected from timber production (e.g., State and National Parks and federal Wilderness). The greatest difference in sample population between old and new inventories was the large National Parks in California and Washington that were measured in the new inventory. Some inventories relocated subplots from the fixed

design if a plot happened to straddle more than one condition class. Condition class classifies variation in a sampled area with respect to land use, forest type, and stand size class. Important differences between inventories at the plot level include the area sampled and the sampling methods used; the new design relies on a standard set of nested, fixed-radius plots centered on four points for sampling trees of different sizes (See Appendix A, Figure A1). However, the old inventory often used a variable-radius method to sample most trees >12.7 cm DBH. This was done using a wedge prism projecting a fixed angle from a central point (Bitterlich 1948, Grosenbaugh 1952). The prism angle (or basal area factor, BAF) varied among regions to maximize efficiency (Table 2; also see Appendix A; Fig. A1), and this affected the probability of sampling trees of different diameters. The probability of tree capture is proportional to tree basal area for the variable-radius method (Grosenbaugh and Stover 1957); whereas, the probability of tree capture is proportional to tree density for fixed-radius plots (Grosenbaugh and Stover 1957). Also, the subplots of the new design, in general, are closer together and fewer compared to the old design, thus capturing less within-stand variation. The tree tally size criteria differ among and within designs (Table 2; Fig. A1 of Appendix A).

The data we used from the old design sampled 100% of the plots on a 1 to 4 year cycle but not synchronously among regions (Hiserote and Waddell 2003). We used data from the new design that sampled 10% of plots across regions per year with a full round occurring every decade. Data from the new design spanned 2001 to 2007, and the data from the old design spanned 1988 to 2000 (with the exception of one ownership, which ended in 2001).

Study Area and Data Preparation. The area of study comprised the conterminous states within the Pacific Coast unit for the FIA program. The included states, California, Oregon, and Washington, are topographically diverse with numerous mountain ranges. Maritime influence combines with complex orographic effects and a wide span of latitude to create a variety of different climatic zones and vegetation types. The region has the broadest range of average annual precipitation (from 25 to 4600 mm/yr) found in the lower 48 United States.

Plot grid density and sample size were equalized among model-building data sets. Plots were selected from the old design that were on the shared base grid with 5.5 km spacing, except for the R5 inventory where this was not possible, which resulted in a subset of plots with mean spacing of 3.9 km. Plots with a tree species recorded more than 100 kilometers outside the species' range (as determined by existing flora and current herbarium records) were examined and removed from the analysis because they were presumed to be errors in identification. A total of 27 plots from the old design and 36 from the new design were removed. Only plots that were at least 50% forested were selected from each dataset, resulting in pools for the old and new designs of 10,831 and 6,950 plots respectively. The larger sample from the old design was randomly sampled to obtain a sample size equivalent to that of the new design (N=6950).

Each data set (one for the old and one for the new inventory) was randomly sampled without replacement to split into two halves of equal size, referred to as the "training" and "testing" data sets. Species' occurrences were summed for the training data. Species with more than 25 occurrences in both training sets were retained for a total of 41 species (Table 3). The training data were used to build models and the testing data were used for model selection and evaluation. Figure 1 shows geographic comparisons of training data set locations among designs.

We extracted climate data corresponding to FIA plot locations to serve as climatic variables or predictors for our species-climate models. We started with 11 variables derived by Ohmann and Gregory (2002) from Daymet grids of the western United States (Thornton et al. 1997) (Table 4). We reduced the number of grids or climate variables as input for our analyses by performing a Principal Components Analysis (PCA) (PC-ORD version 5.2; McCune and Mefford 2006). We first gathered data for the PCA by taking a random sample comprising 8,000 points across Washington, Oregon, and California. We extracted the corresponding data from the 11 climate grids. We performed a PCA based on a matrix of correlation coefficients among the data. We selected the first four components that represented 97% of variability in the data (Table 5). The four corresponding eigenvectors were then used to generate the new grids representing PCA

scores across Washington, Oregon, and California (Table 5). The use of PCA scores as predictors in our species-climate models ensured statistical independence among the predictors. The PCA scores also simplified the comparison among inventories by reducing the number of predictors.

Model building. Non-parametric multiplicative regression (NPMR) was used to build the climate niche models using Hyperniche 2.0 (McCune 2006, McCune and Mefford 2008). This technique was chosen among numerous empirical techniques in the literature for species-habitat models (e.g. Elith et al. 2006, Guisan et al. 2007) as it captures the nature of biological response to multiple interacting factors (McCune 2006). Further, current theory supports that species' response patterns take non-linear, complex shapes (Austin 2002, Oksanen and Minchin 2002), and when tested against other popular techniques for modeling different shapes of simulated data sets, NPMR proved to be most tractable with any shape of underlying data structure (Lintz et al. 2010).

NPMR objectively optimized kernel width by maximizing fit. To do this, a local window centered on a target point within the predictor space (as with other kernel smoothers). A weight function was applied within the window to assign numeric weight to points surrounding the target that decrease with distance from the target. A local model (the local mean) predicted the dependent variable at the target point. NPMR repeated this procedure for all target points to generate a prediction curve or surface. NPMR omitted the target point when predicting the response at that point and multiplied weights from individual predictors to automatically accommodate complex interactions in many dimensions.

To adequately gauge the effect of inventory on species-climate models, we compared statistics from models built from the different data. We also compared qualitative attributes of the models and geographic maps of predicted probabilities of occurrence generated from the different models. We juxtaposed comparisons of the old and new inventories (referred to as 'old-new') with comparisons of the two subsamples of the new inventory (referred to as 'new-new'). The old-new comparison used the training sets from each inventory for model building. The new-new comparison used both

the testing and training data sets within the new design for model building. Each half of the data set from the new inventory was a sample for model building with the other half used for model testing. In this way, we examined the type of error that can arise in NPMR from taking different samples of the same data set. The data sets were so large that bootstrapped confidence bands were both unnecessary and too time consuming. We reasoned that if the same magnitude and type of deviations from the predicted 1:1 line across species in the 'old-new' comparison are also seen in the 'new-new' comparison, then error arising from sources other than sample design or environmental change were likely the cause.

Model Evaluation and Selection. Two measures of model performance for binary data were used for different purposes, model selection and model evaluation. First, we used the $\overline{LogB}$ or the average contribution of a sample unit (or FIA plot) to the log likelihood ratio for model selection (McCune 2006). Popular statistics for model selection with classifiers are often derived from the likelihood ratio such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) (Hastie and Tibshirani 2001); however, the AIC and BIC approximate the optimization curve used in model selection for sample sizes too small to generate the curve empirically (Hastie and Tibshirani 2001). The empirical optimization curve represents the loss in externally-validated fit that occurs as model complexity increases. Theoretically, the model with the greatest externally-validated likelihood ratio is the most robust and accurate. Given our large sample size, instead of relying on an index to approximate this curve such as an AIC, we rely on external validation itself using $\overline{LogB}$. This is considered to be the ideal scenario for model selection that can occur only when sufficient data are available (Hastie and Tibshirani 2001).

Second, we use the Area under the Receiver Operator Characteristic (ROC) curve or AUC for model evaluation (Hanley and McNeil 1982). The $\overline{LogB}$ is not suited to comparing the fit of models among species with different prevalence. Instead, the AUC is best for comparing model performance among species that differ in prevalence as it does not depend on prevalence (Fawcett 2006). Hence, we used the externally-validated AUC

to compare the best models among species. An AUC of 0.5 represents a model fit no better achieved by chance alone. The maximum value of the AUC is 1.

Climate and species' occurrence data from plots withheld from model building were used to calculate externally-validated measures of model performance. The withheld climate data were supplied as input to the candidate models (after the models were built from the training data sets), and the corresponding withheld species' presence/absence data were compared to the resulting predictions.

The relationship between the externally-validated AUC and the externally-validated $\overline{LogB}$ showed strong correlation across species for the best candidate models (See Appendix A; Fig. A2). Each species displayed a different shape and slope of the relation between the AUC and $\overline{LogB}$due to the dependence of $\overline{LogB}$on species' prevalence. Model performance was linked to the number of predictors (Fig. A2).

Sample Design Effect on Models and Data. We compared models based on the two sample designs using the following measures: the fit as externally-validated AUC, the quality of the predictors or the type of predictors chosen for a model, random deviations from the 1:1 line for predicted values among models of the same species, and systematic deviations from the 1:1 line (or whether models from one data set tended to overestimate or underestimate probability of occurrence, relative to the other data set). Model comparisons were generated by obtaining predictions from models based on the same random sample of 3,475 points within the study area. Predictions were compared for old-new and new-new comparisons. The random error or non-directional deviation from the 1:1 line was derived using Root Mean Squared Prediction Error (RMSE). However, the RMSE is a function of the maximum in predicted probability of occurrence for a species, which varied with species' prevalence. Hence, the RMSE was standardized to compare across species with different maxima in probability of occurrence. The differences from the 1:1 line were divided by the range of the data or maximum probability of occurrence before squaring and summing to yield the 'normalized RMSE' or NRMSE. The NRMSE is often expressed as a percentage. The NRMSE represents the degree of deviation from the 1:1 line as a percentage of the axis length. The Wilcoxon

signed-rank test evaluated the null hypothesis that the median in NRMSEs from the new-new and old-new comparisons were equal (Wilcoxon 1945). The systematic deviations from the 1:1 line were also compared. The residuals were standardized by their range, and three quantiles were plotted and compared for the old-new and new-new comparison, the 25^{th}, 50^{th}, and 75^{th}. These indicate the central tendency of the residuals and whether one data set tends to model and predict a greater probability of occurrence compared to another. The median standardized residual tracks the median *non-zero* standardized residual closely and linearly except for values very near to zero, which were not meaningful.

Raw differences among data sets were investigated to aid the interpretation of modeled comparisons. We compared several metrics from the data: 'climatic bias,' probability of tree capture, and species' prevalence. Climatic bias was defined as the disagreement among two histograms where each histogram represents the frequency of climatic data values corresponding to locations where a species' was found (Kadmon et al. 2003). This characterized and compared species-specific structure of climatic data from two different samples of presence/absence data. Although, instead of calculating a statistic to characterize climate bias, we examined it visually for reasons discussed in Appendix B. Also, we derived gross measures of probability of tree capture across sample designs using the average number of trees per FIA plot by management region. Before the switch from the old to new design, approaches to sampling differed in numerous aspects by management region; hence, we used management regions as sample units to examine probability of tree capture across designs.

Sample Design Effect on Maps. We used geographic grids of PCA scores of climate variables across the study area as input to make maps of probability of species' occurrence for species-climate models (using Hyperniche 2.0 and ArcGis 9.2). We mapped differences among predictions for new-new and old-new comparisons. This was done for three species, *Arbutus menziesii*, *Tsuga heterophylla*, and *Pinus ponderosa*. The species were chosen to span a representative range of prevalence in the data (N=301, 613, 958 respectively). We extrapolated conservatively or little beyond the existing range of

the data using the default setting in Hyperniche 2.0 where the minimum neighborhood size (in environmental space) for an estimate (in geographic space) is equal to or greater than a quarter of the average neighborhood size for a model. The average neighborhood size is the average amount of data bearing on the estimate of the response variable at each point.

RESULTS

Sample Design Effect on Models. Most species from the old and the new FIA sample designs yielded similar predicted values for models of species' probability of occurrence (models built from the different data sets were given the same new set of climatic data as input) (Fig. 2). The mean NRMSE in the old-new comparison was 6% (95% quantiles: 3%, 12%) compared to a mean of 4% (1%, 10%) in the new-new comparison for the 41 species. The difference among median NRMSEs (across species for old-new and new-new comparisons) was not likely to be due to chance alone (Wilcoxon signed-rank two-tailed, $p < 0.001$). Additionally, many species that showed strong systematic deviation from the 1:1 line among designs (detectable by eye) also showed this type of deviation for models built from different samples of the same design (AECA, CHNO, PIAL, POBAT, QUAG, QUDO, QUWI; Fig. 2, Fig. 3). Others showed greater deviation from the 1:1 line in the old-new comparison than in the new-new (ABAM, ABMA, CONU, PIJE, PILA, PIMO, PIMONT, PIPO, PISA, QUKE, SESE). Many of these species comprised a subset of 16 species with prevalence below 100 among training data sets (where N=3475; Table 3).

The median standardized residuals gauged systematic error in the residuals and remained close to zero for most species. Medians ranged from 0.02 to -0.06 among old-new and new-new comparisons (Fig. 4). Four species, ABCO, CADE, PIPO, and PSME, had greater probabilities of occurrence predicted for the old design compared to the new (Fig. 4). The medians of the deviations were, however, small (absolute systematic deviation < 0.06). Systematic error was present but substantially weaker in the new-new comparison than in the old-new comparison for CADE, PIPO, and PSME (Fig. 4).

The differences among models built with old and new data did not appear to be caused by model selection of different predictor variables. The same number and type of predictors were chosen among models for 26 species in the new-new and 26 species in the old-new comparisons (see species codes with asterisks in Figues 2 and 3). Most models selected four predictors out of the four available (an overall mean of 26 species among models built from different data pools). Only four instances occurred with models containing two predictors, and the rest contained three.

All models contained the first PCA component. Also, most predictions fell near the 1:1 line even for models without the same number and/or type of predictors (e.g. ABAM, ABGR, ABLA in Fig. 2 and 3). Despite the differences among predictors for many of the compared models, agreement in model fit for old-new was strong and similar for new-new (NRMSE of AUC=8% for new-new and 13% for old-new; Fig. 5A, Fig. 5B). The minimum model fits (or the externally-validated AUC) among the comparisons were 0.766 (new-new) and 0.835 (old-new), and the maximum models fits were essentially the same, 0.995 and 0.996 (new-new and old-new respectively). The mean externally-validated AUC across all models was 0.935. Four species consistently had fits above 0.975 across data sets: *Pinus monophylla*, *Sequoia sempervirens*, *Quercus douglasii*, *Quercus agrifolia*. These species have prevalence under 100 and occur mostly in California. Most species' with strong fits tended toward increased agreement among predictions (Fig. 5C, Fig. 5D). However, the agreement between compared models in type and number of predictors did not play a role in these relationships (see symbol coding in all subplots of Fig. 5). Species with high prevalence had slightly lower fit than species with low prevalence, albeit very weakly (nonparametric regression, cross-validated r^2= 0.08) and nonlinearly (results not shown). Model fit did not vary with sample design across all models.

Sample Design Effect on Data. The mean number of trees per plot by size class (where the mean is used as a proxy for probability of tree capture) deviated strongly from the expected 1:1 line (RMSE=8 or NRMSE=25%). The deviation was pronounced for all but the largest trees (Fig. 6). The number of trees per plot were higher for the R6 and

BLM areas for all tree sizes in the old design compared to the new design. For the regions using variable-radius sampling in the old inventory, numbers of small trees (12.7-38.1cm dbh) per plot were lower in the old inventory, while numbers of larger trees (38.1-76.2 cm) were higher in the old inventory (Fig. 6; tree selection probabilities for each design are shown in Figure A1).

Differences among the ECDFs (empirical cumulative distribution functions) from different sample designs for each species was more visually pronounced for species with prevalence <100 (e.g. CONU, PIMO, QUAG, and POBAT) (Fig. 7). However, the difference among two ECDFs drawn from the same population is always more pronounced for smaller samples (see Appendix A; Fig. A4). Still, little disagreement or climatic bias was evident for most species (where climate bias is the difference among two ECDFs of climate values corresponding to species' locations) (Fig. 7). One would expect that species with the most pronounced systematic error in the modeled predictions would demonstrate greatest evidence of climatic bias in the data (which may be attributed to sampling differences). This was not clearly the case as ABCO, CADE, PIPO, and PSME did not have climatic bias greater than that shown for species without systematic error among predictions (e.g. ALRU, QUCH, TSHE, LAOC, CHNO, ABSH)(Fig. 4, Fig. 7). However, for species with greater prevalence like PSME, a small gap likely has more ecological consequence and meaning compared to the same gap in a species with low prevalence (see Fig. A4). To zoom in on ABCO, CADE, PIPO, and PSME we used Quantile-Quantile plots (QQ-plots). QQ-plots among two samples will be linear if two samples come from the same distribution. The ABCO QQ-plots showed the strongest exception to linearity for PCA2, and the tail of the QQ-plot for the old-new comparison deviated compared to the new-new comparison (Fig. 8).

Sample Design Effect on Maps. Coarse patterns in predicted probability of occurrence in geographic space appeared similar between sample designs within species, except for the aerial extent where predictions were made (Fig. 9). Also, maps of the differences among predictions within the old-new and new-new comparisons revealed spatial clustering of strong differences at different scales (Fig. 10). These patterns (e.g. in the

highest magnitude residuals per map) tended to follow broad gradients in topography and climate rather than cluster by management region (Fig. 10). A map of FIA management regions in the old inventory is provided (Appendix A; Fig. A5). Spatial clustering does not appear to correspond with these regions. Spatial clustering at scales much smaller then FIA management units was evident across new-new and old-new comparisons (e.g. the contiguous blue patch in the upper left portion of the map of "ARME new-new," Fig. 10). The source of that error is unknown but is likely due a contributing variable not included in models such as fire or competitive exclusion.

The maximum absolute differences in predictions differed among species (Fig. 10). These magnitudes depended on the difference among predictions for a location and the range in probability of occurrence predicted for a species. In all three examples the old-new had the highest absolute differences among the comparisons.

DISCUSSION

We found two likely effects of inventory method on niche models and their predictions. First, there is a 2% increase in random error among modeled predictions when using one design to predict occurrences in the other design (an average 2% more than within-design sample error for a single within-design comparison). Second, small quantifiable systematic error is present for 4 out of 41 species, yet the error for only one species, ABCO, shows the strongest evidence for a link to inventory. Each of these two effects are discussed below.

Since the 2% effect in NRMSE was evident as as a central trend across species, systematic environmental change such as climate change is probably not the main reason. This argument is based on evidence from the past as climate change affected tree species' ranges in the northern hemisphere differentially rather than uniformly during periods of historical migrations (Davis and Shaw 2001). However, it is not possible to completely rule out climate change or other potential causes that differ from inventory type. The forest at time t can be slightly different than the forest at time t + 10 years for natural reasons such as logging, fire, pathogen outbreak, or recruitment patterns that follow

changes in climate. However, presence/absence is a coarse-scale variable, and trees are long-lived. For such natural factors to translate to presence/absence requires a large magnitude of effect.

We believe the effect in NRMSE for the old-new comparison is due largely to the change in inventory methods for the following reason: probability of occurrence is a function of plot size, and plot sizes differed across management regions in the old inventory. Larger plots were used for all tree sizes on R6 and BLM lands while other management regions used variable radius plots with plot size proportional to tree diameter. At least with respect to tree fequency, the varied effects of different regional inventories on tree frequency are suggested (Fig. 6). Hence, when the data from the old design were pooled across regions with different protocols, the cumulative effect (of different types of bias imposed by different sampling methods) likely manifested as random error that propagated to modeled predictions. The difference in plot size that changes a presence record to an absence record at a local scale will depend on the species' density and size distribution for each of species. Our work suggests that the variation in size of FIA plots within and among inventories probably has a small effect on presence/absence data patterns.

Inventory change does not seem to be the cause for the systematic error associated with *Calocedrus decurrens*, *Pseudotsuga menziesii*, and *Pinus ponderosa*. We reach this conclusion given the assumption that the effect of the inventory would first manifest as climatic bias (where climate bias is the difference between histograms of climatic data corresponding to presence locations), which was not clearly the case for these species. Also, systematic error among modeled predictions was present for these three species in the new-new comparison (albeit weaker than the systematic error for the old-new comparison).

Conversely, evidence of climatic bias existed for *Abies concolor* (Fig. 4, Fig. 7, Fig. 8). *Abies concolor* is part of an intergrading complex of species with *Abies grandis* (Critchfield 1988). Field discernment of the two species can be difficult where they hybridize yet the species are ecologically distinct (Ferrell and Smith 1976, Zobel 1974).

Protocols for identification of both species changed in some areas from the old to new design. Old data from R6 national forests tended to either use one species code or the other in areas where new crews have used both, which would affect the tails of the frequency distributions in climate (see Figures 7 and 8). This change of protocol associated with inventory shows up as climatic bias because it changes the geographic distribution of presences. The change manifests as difference between frequency distributions of a climate variable.

Spatial patterns in the mapped differences between predictions occurred across scales for the old-new as well as the new-new comparisons. To further examine whether differences among inventories may be the cause for these patterns, we clipped the maps shown in Figure 10 by management regions (see Appendix A; Fig. A5) and compared distributions of the differences for old-new and new-new by region. Although the disparity in median differences (among mapped predictions) was greater for select management regions in old-new compared to new-new, the patterns did not correspond clearly with differences in inventory (not shown). Apparently the climate predictors did not segregate by management region enough to affect the distribution of residuals within them. The overlap in climate values among regions with different plot designs probably explains why, despite greater error, the accuracy of the species' predictions were high and similar among designs (e.g. revealed by externally-validated AUCs).

Spatial autocorrelation. Maps of the differences between predictions (Fig. 10) show spatial autocorrelation. Spatial autocorrelation occurs when observations in close spatial proximity tend to be more similar than expected for observations more spatially separated (Schabenberger and Gotway 2005). The treatment of spatial autocorrelation in niche modeling is a topic of active research (Segurado et al. 2006, Miller et al. 2007, Dormann et al. 2007, Fortin and Dale 2009). In our work, we used a kernel smoother in climate space. The climate variables were spatially auto-correlated in geographic space. We ignored the autocorrelation structure among the data in climate space, and this is known as the "working independence" method in kernel regression, which has theoretical basis (Ruckstuhl et al. 2000, Lin and Carroll 2000). Kernel smoothers are designed to

accommodate spatially dependent data (e.g. Yakowitz and Szidarovszky 1985, Wahba 1990, Hutchinson and Gesler 2004). Spatial aggregation at variable scales should be expected in the maps of differences among predictions (our results). We do not attempt to account for local processes that can govern spatial autocorrelation (e.g. in local climate, which potentially explains much variability in species' presence/absence). Additionally, spatial aggregation in residuals can result from failing to account for other potential drivers.

Cross-validation methods. We found noteworthy variation in modeled output due to within-design sub-sampling. This variability can be mitigated through a different method of model selection. Model selection by external validation (the method we used here) is a variation of *k*-fold cross-validation where $k = 2$. Model selection with greater than 2-fold cross-validation uses models built from repeated sub-sampling of the data (henceforth called cross-validation). In cross-validation, the sample is randomly divided into $k > 2$ subsamples. One sub-sample is retained (out of the *k* sub-samples) as the testing or validation data, and the other k-1 subsamples are used each to train or build models. Results of model performance for iteration through *k* subsamples are combined for optimization. The model that is most robust to external data and accurate is selected.

We further examined the reason behind differences in the quality of models for the subsets of models choosing different predictors for a species and among sub-samples. The results presented here suggest that model selection using cross-validation with NPMR may result in greater precision (without compromising robustness) compared to external validation. This interpretation is based on preliminary results (not shown) stimulated from the modeling results of this work. New-new predictions generated from models selected by external validation were compared with new-new predictions generated from models selected by cross-validation. The agreement among predictions improved and model similarity increased for model selection using cross-validation. This is an interesting (albeit preliminary) finding given that the statistical paradigm considers external validation as the gold standard in model selection (Hastie and Tibshirani 2001).

More research is needed in this regard with respect to kernel regression, NPMR, and large sample sizes.

Sampling effects. Sampling has been identified as an area requiring further investigation in niche modeling especially given the haphazard nature of some data sets (Elith et al. 2006, Guisan and Zimmerman 2000, Araújo and Guisan et al. 2006, Heikkinen et al. 2006, Hampe 2004). Recent papers that address sampling issues with respect to niche modeling emphasize presence-only data and explore sample size (e.g. Pearce and Ferrier 2000, Stockwell and Peterson 2002, Elith et al. 2006, Hernandez et al. 2006), sample completeness (e.g. Kadmon et al. 2003), sample attributes for optimal model validation (e.g. Araújo et al 2005), idealized sampling strategy for niche models (e.g. Wessels et al. 1998, Reese et al. 2005), and sample bias due to site accessibility and/or presence-only data (Reese et al. 2005, Kadmon et al. 2003, Kadmon et al. 2004). Sampling bias can add extraneous error to ecological signals especially with presence-only data, which can mislead model development, spuriously increase or decrease fit, and affect spatial predictions (Barry and Elith 2006, Guisan et al. 2006, Araújo and Gusian 2006). Our work suggests a small effect occurs on predictions from niche models due to a major change in a large biological inventory, which, in part, represents a change in sample design.

CONCLUSION

The many features that changed with the overhaul of sample design for the Forest Inventory Analysis Program had a small cumulative impact on niche models and maps of probability of occurrence based on tree species' presence/absence data. Further, the fit of all the models was high across data sets. Tree species' occurrence is strongly linked to climate in the Pacific coastal United States. Modeled predictions in probability of occurrence disagreed little between inventories and also between two sequential decades. However, presence/absence data and their predictions will not necessarily detect low-level mortality or changes by life stages. Finally, our ability to describe and probe the basis of species' distributions depends on the quality of our data. This work corroborates

with the pervasive and pressing need to quantify different types of error in niche modeling to address issues associated with large-scale data integration (Barry and Elith 2006, Elith et al. 2002).

ACKOWLEDGEMENTS

We thank the U.S.D.A. Forest Service for funding this work.

FIGURES

Figure 1. Study area. Locations of FIA plots for training samples (N=3475) are shown for the old and new design.

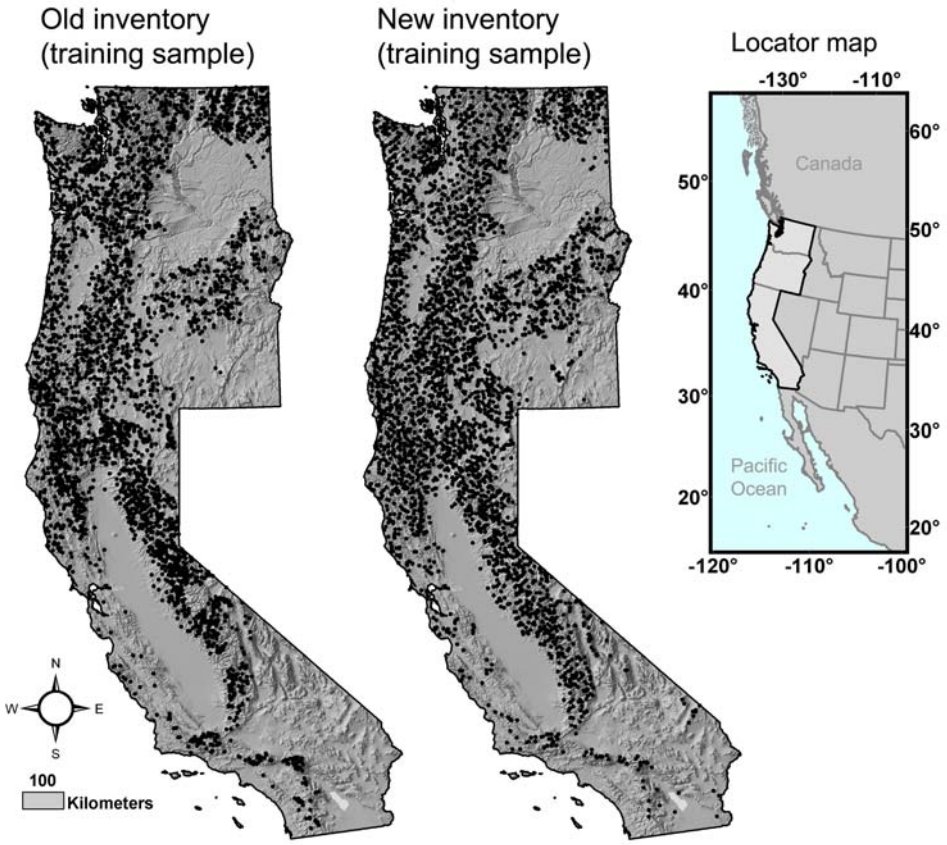


Figure 2. Predicted probability of occurrence by species for models built from new versus old sample designs. Red line represents the ideal 1:1 line. Predicted values were generated from a random sample of unseen climate data (N=3475). Species codes with asterisks* denote that the compared models for that set of axes had the same model functional form or number and type of predictors.

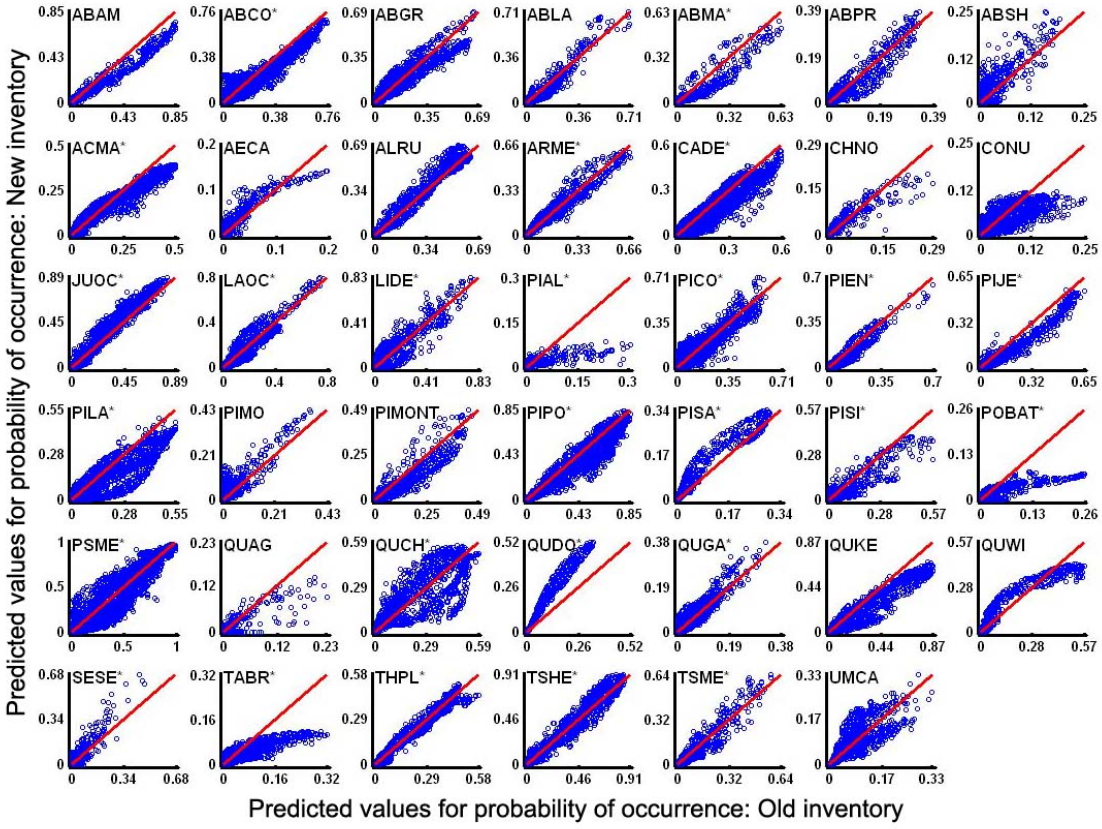


Figure 3. Predicted probability of occurrence by species for models built from 'new versus new' sample designs where a random sample from one design is compared to a second random sample from the same design. Red line represents the ideal 1:1 line. Predicted values were generated from a random sample of unseen climate data (N=3475). Species' codes with asterisks* denote that the compared models for that set of axes had the same model functional form or number and type of predictors.

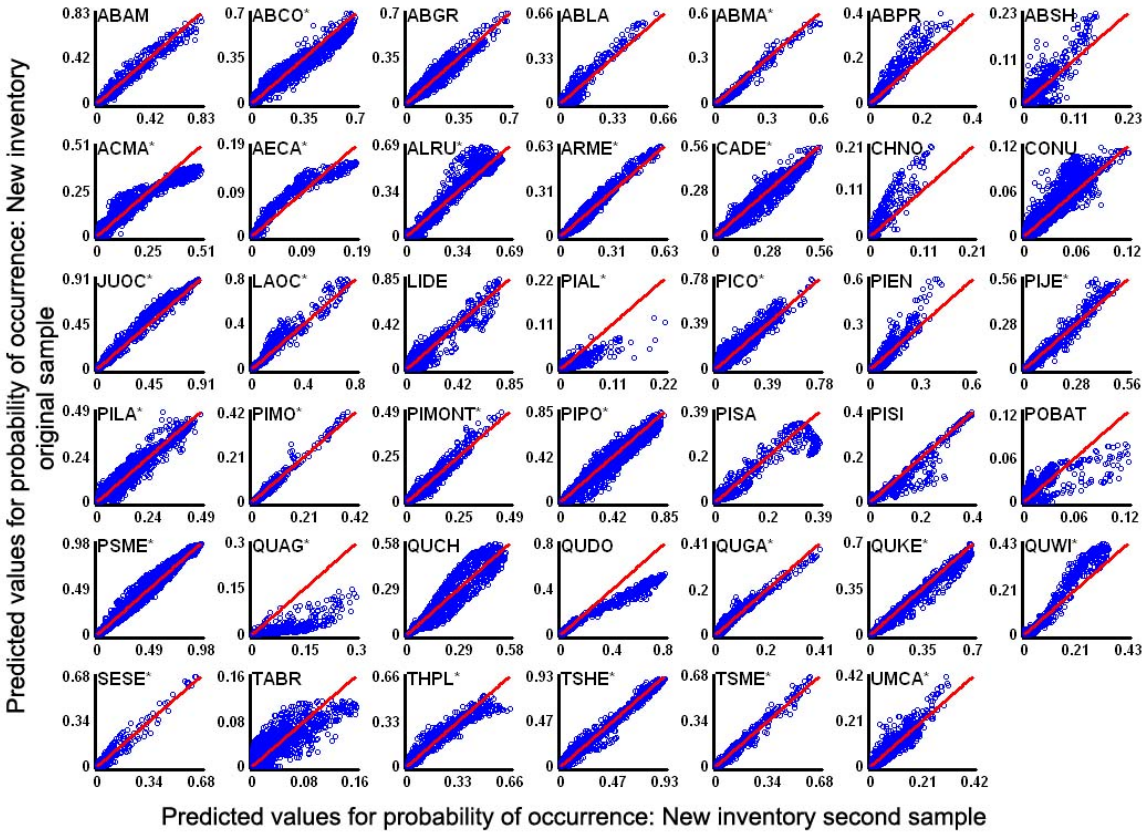


Figure 4. Stem plots show the central trend in the standardized residuals with respect to the 1:1 lines in Figures 2 and 3. The mean (white dot), the 25th quantile (black dot), and the 75th quantile (black dot), are shown by species (rows) and by comparison, old-new (left) and new-new (right).

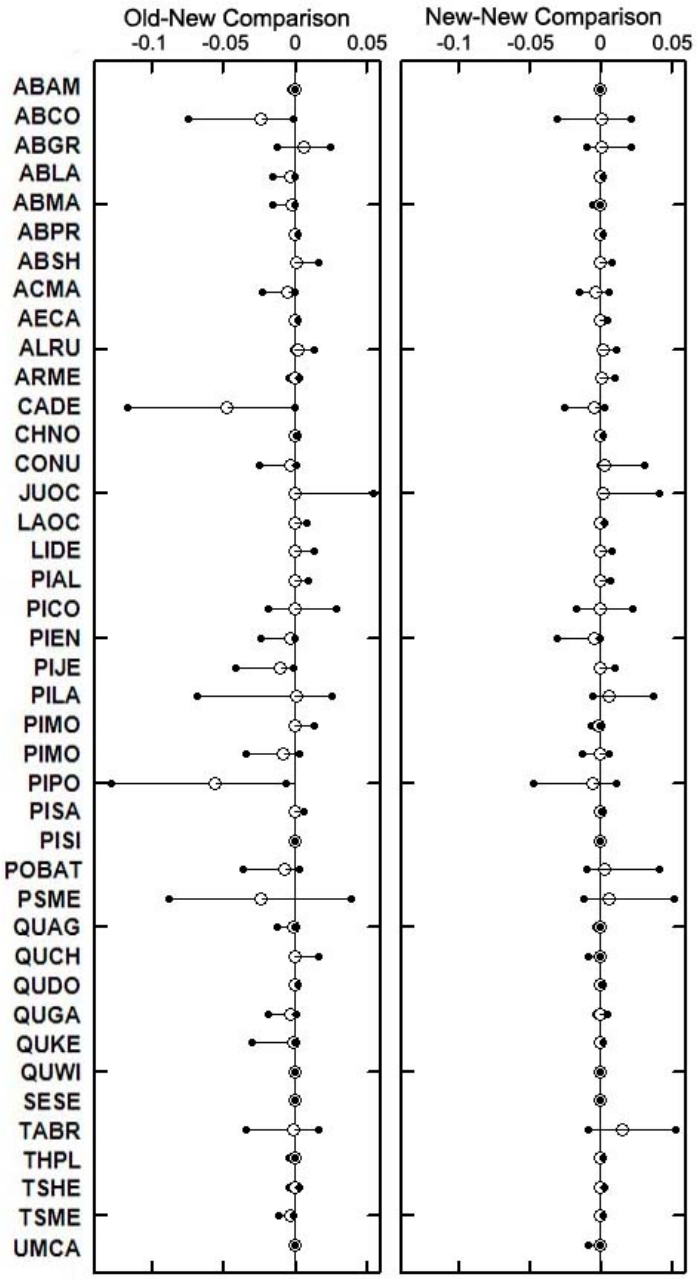


Figure 5. Top row: Externally-validated area under the receiver-operator characteristic curve (AUC) compared from old-new comparison and new-new comparison. Bottom row: Standardized RMSE (NRMSE) versus mean AUC among models for old-new and new-new comparisons. Points represent species. Paired models where qualitatively different predictors were chosen for the same species are shown with circles. Paired models where the same predictors were chosen for a species are shown with triangles.

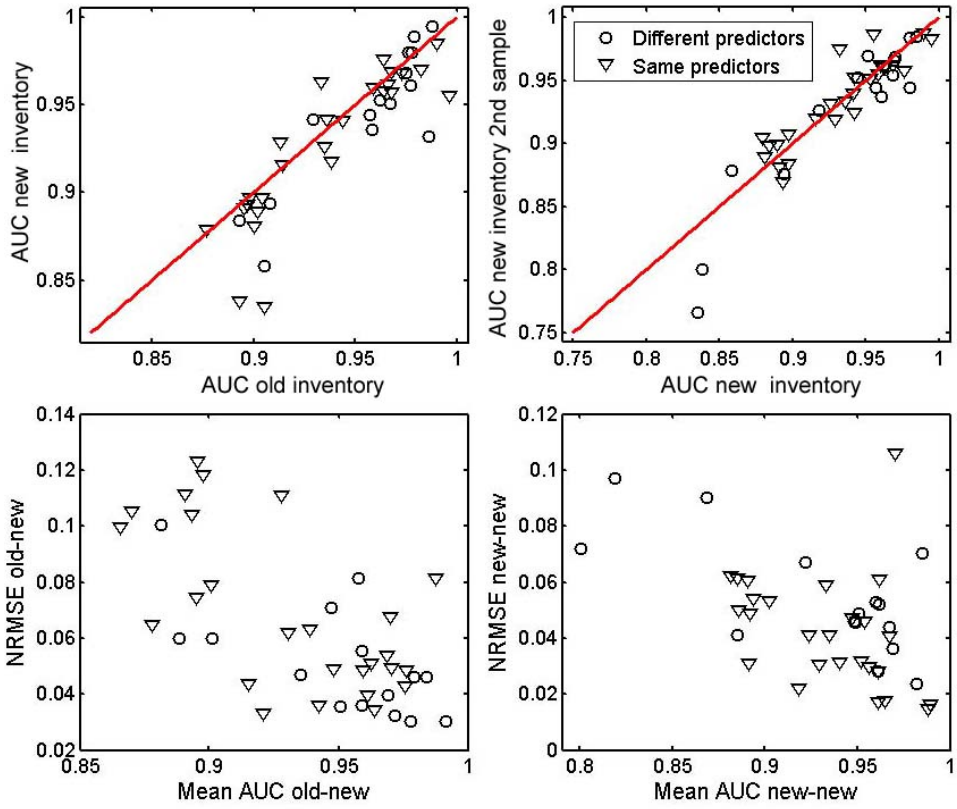


Figure 6. Mean number of trees per plot and by age class. The ideal 1:1 line is shown. Each point represents a size class in a particular management region from the old design (see Table 1).

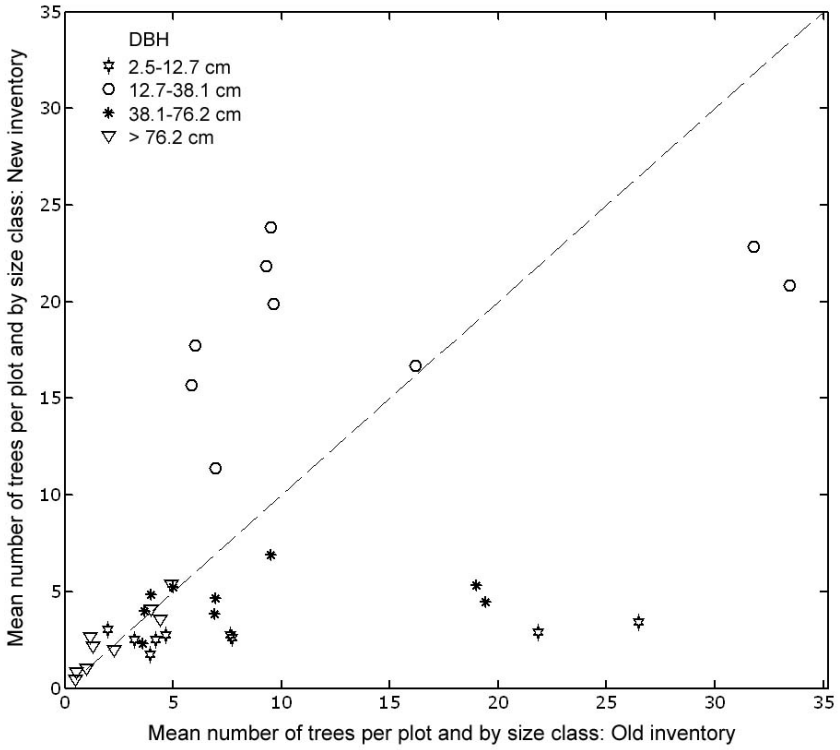


Figure 7. The ECDF (empirical cumulative distribution function) for each climate variable (or PCA component) as it corresponds to species' occurrence for each design (old and new). Climate variables are color coded, magenta shows PCA1 from both designs, black shows PCA2 from both designs, blue shows PCA3 from both designs, and red shows PCA4 from both designs.

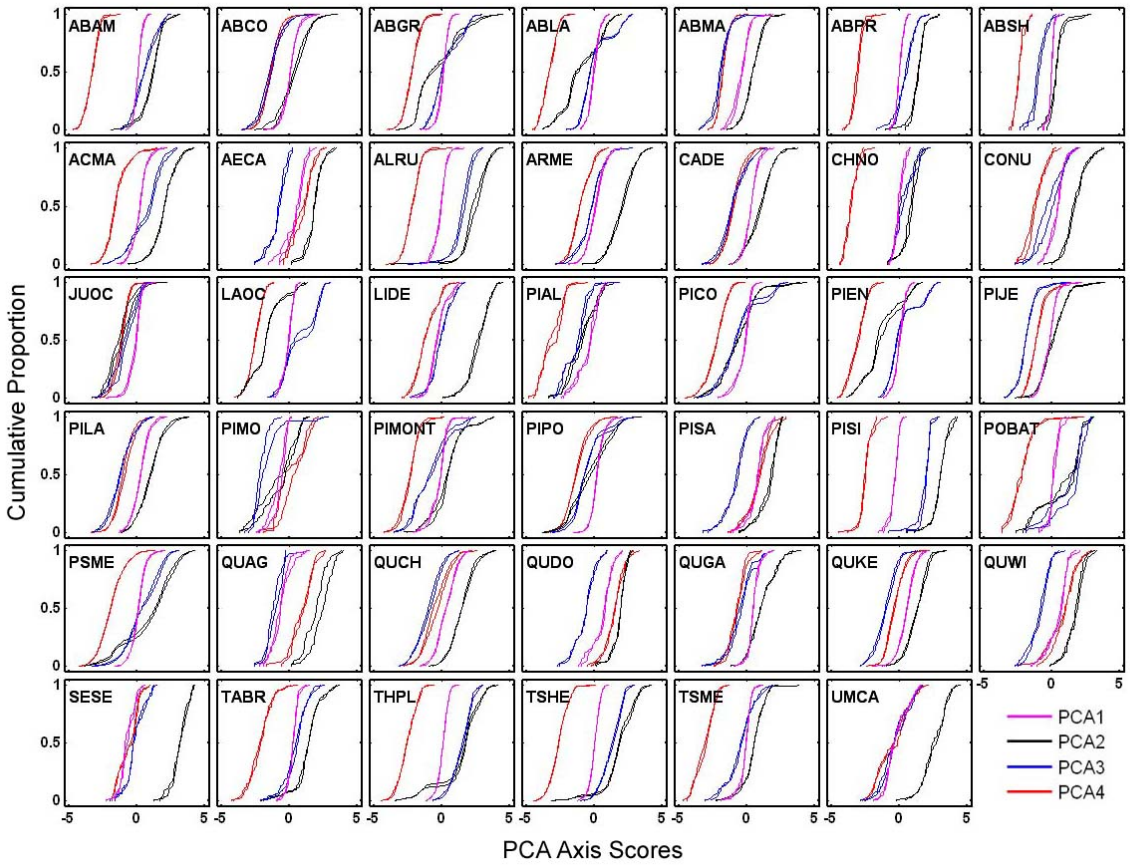


Figure 8. QQ-plots are shown comparing the frequency distributions of climate variables corresponding to species' presence records across four species (rows) and climate variables (columns). QQ-plots for the old-new comparisons are shown in black and QQ-plots from the new-new comparisons are overlaid in pink.

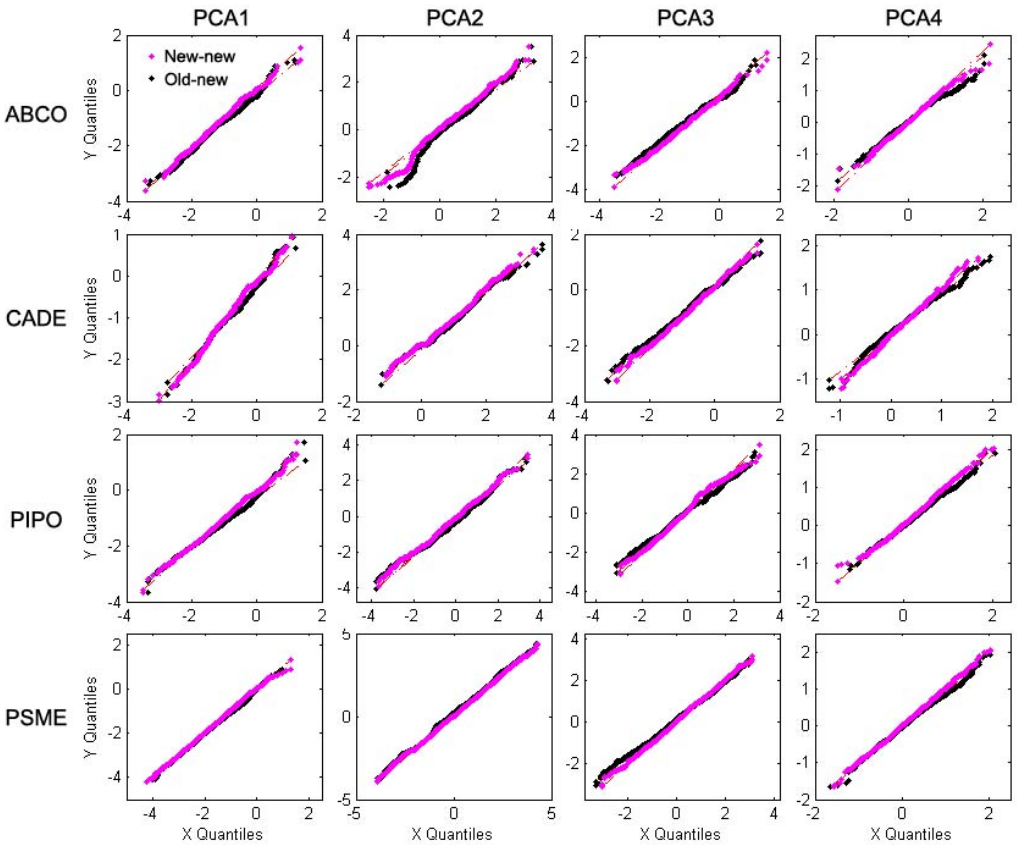


Figure 9. Maps of probability of occurrence are shown for three species *Arbutus menziesii* (top row), *Pinus ponderosa* (middle row), and *Tsuga heterophylla* (bottom row). Maps correspond to prediction from each data set compared (see Fig. 2).

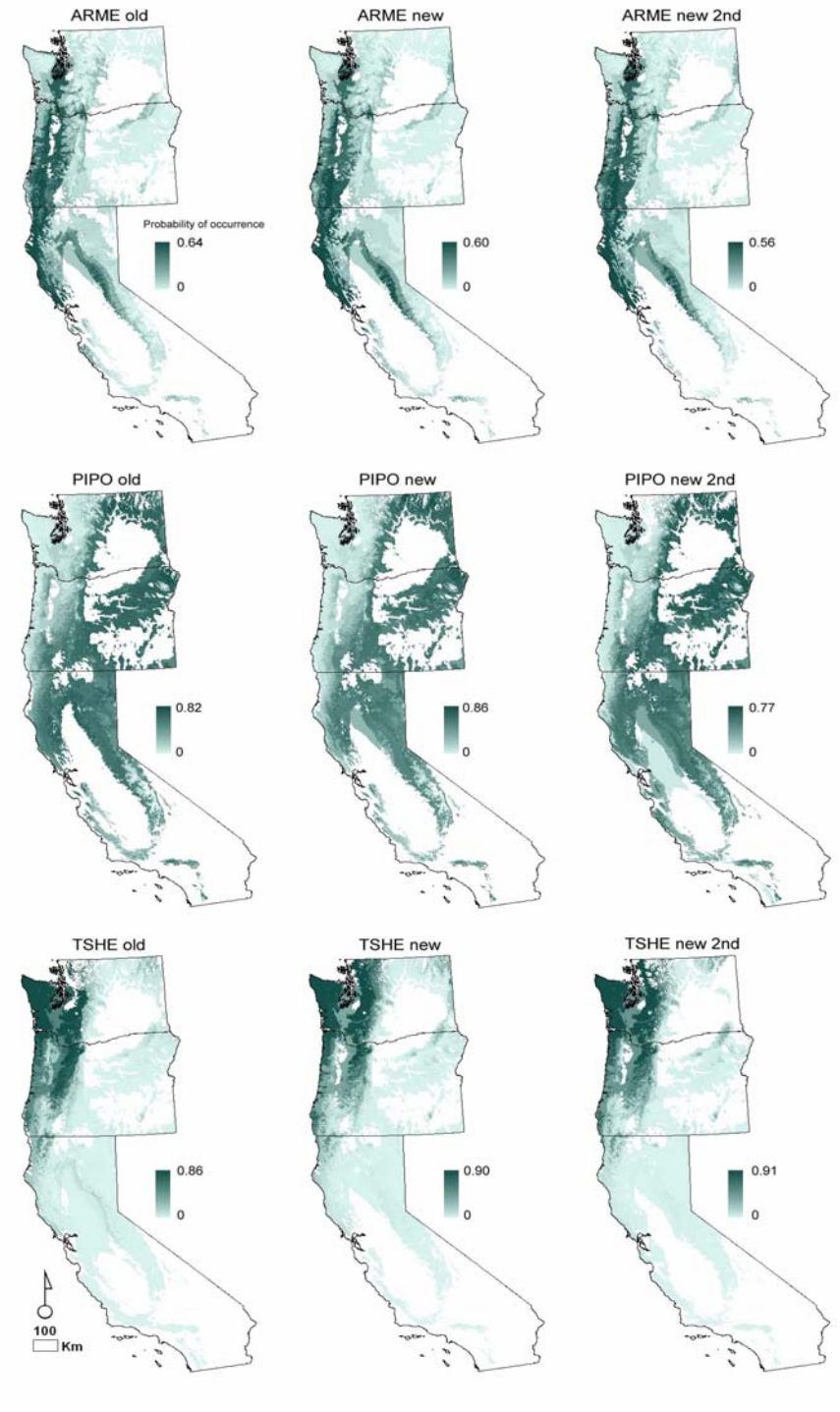
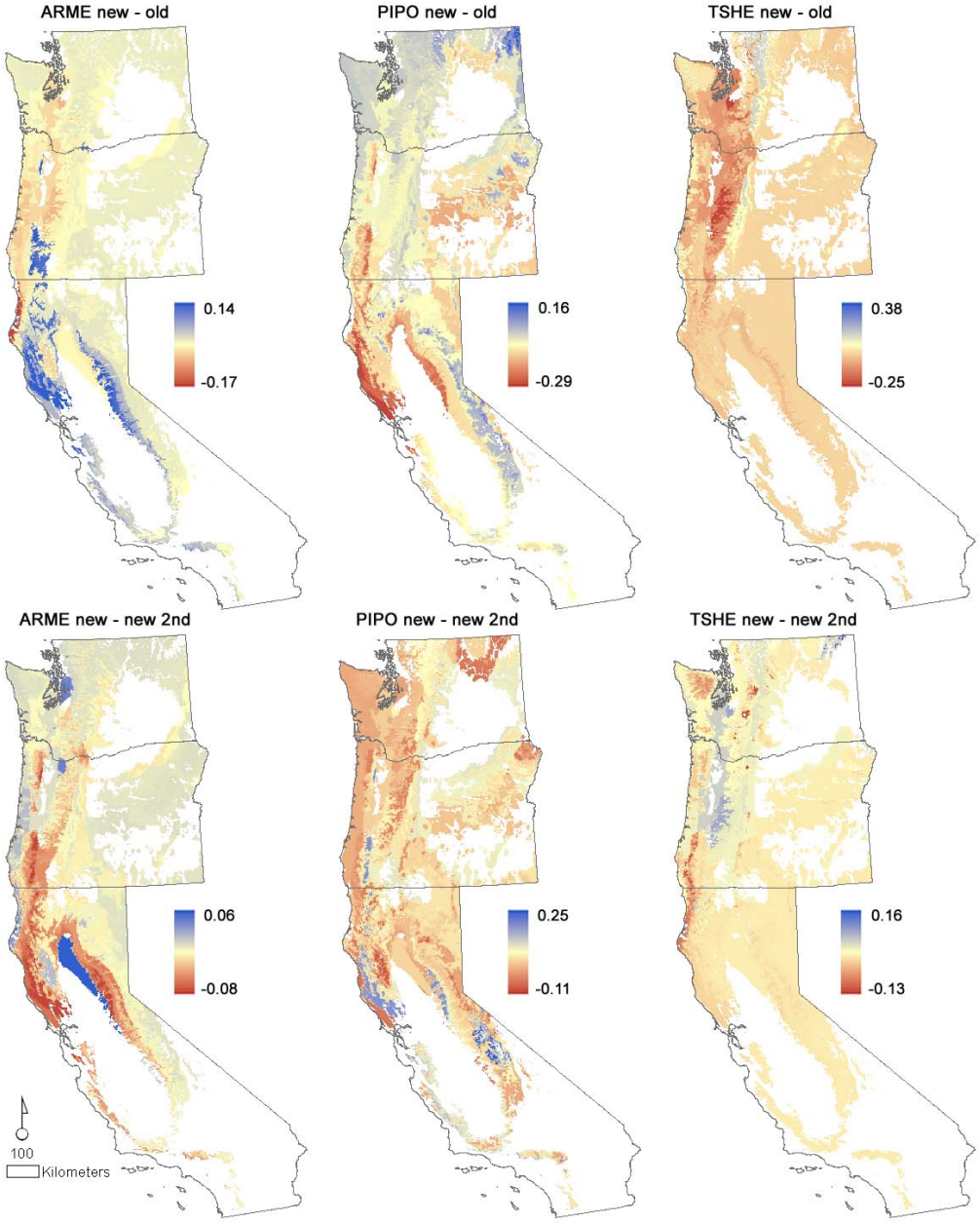


Figure 10. Maps of differences among probabilities of occurrence show regions of uncertainty among the maps of probability of occurrence (see Fig. 9). The difference for the old-new comparison equals the new minus the old (maps shown in Figure 9 were subtracted). The difference for the new-new comparison equals the second sample of the new minus the primary sample of the new. Differences are shown for the old-new comparison (top row) and the new-new comparison (bottom row) for the three species in Figure 9, *Arbutus menziesii* (left column), *Pinus ponderosa* (middle column), and *Tsuga heterophylla* (right column).



TABLES

Table 1. Data source definition, inventory dates, and sampling density by inventory units in California, Oregon, and Washington (Hiserote and Waddell 2003).

Data Code	Source name	States	Dates of Inventory	Distance between points of sample grid
Old, Periodic Inventory				
WWA	FIA, Western Washington	WA	1988-1990	3.9 km
EW	FIA, Eastern Washington	WA	1990-1991	5.5 km
CA	FIA, California	CA	1991-1994	5.5 km (7.7 km in oak woodland)
WOR	FIA, Western Oregon	OR	1995-1997	5.5 km
EOR	FIA, Eastern Oregon	OR	1998-1999	5.5 km
R6	Forest Service, Region 6, Pacific Northwest	OR, WA	1993-1997	2.7 km (5.5 km in wilderness)
R5	Forest Service, Region 5, Pacific Southwest	CA	1993-2000	Numerous (5.5 km base grid)
BLM	Bureau of Land Management	Western OR	1997	5.5 km
RMRS	FIA, Rocky Mountain Research Station	Eastern WA, Eastern CA	2001, 1997	5.5 km
New, Annual Inventory—all states				
PNW Annual	FIA	CA, OR, WA	CA and OR began in 2001;WA in 2002	5 km

Table 2. Differences in plot-level protocols by inventory units for the "large tree" size class. Variable radius plots were sampled using a wedge prism.

Unit code	Plot Radius (m)	BAF* (m^2/ha)	Sub-plots (#)	Total area (m^2)	Tree tally size criteria (DBH, cm)
Periodic Inventory					
WWA	variable	6.9	5	variable	>17.8**
EWA	variable	9.2	5	variable	>12.7
CA	variable	6.9	5	variable	>17.8
WOR	variable	6.9	5	variable	>12.7
EOR	variable	4.6 or 6.9	5	variable	>12.7
R6+BLM	8.016	N/A	5	1009.4	>7.6
R6+BLM	15.575	N/A	5	3810.6	>33
R5	variable	4.6 or 9.2	5	variable	>12.7
Annual Inventory					
PNW Annual	7.32	N/A	4	672.5	>12.7
PNW Annual	15.575	N/A	4	4050.1	>76.2 west; >61 east

*BAF stands for basal area factor.
**DBH stands for tree diameter at breast height.
"West" and "east" refer to locations relative to Cascade Mountains in Oregon and Washington. In California, the same minimum diameter as "east" was used.

Table 3. Forty-one species were studied. Species were retained for study if they contained more than 25 occurrences in the primary training data sets from the old and the new design.

Code	Latin name	Common name	Prev-alence (old)	Prev-alence (new)
ABAM	*Abies amabilis*	Pacific silver fir	243	213
ABCO	*Abies concolor*	white fir	801	579
ABGR	*Abies grandis*	grand fir	333	351
ABLA	*Abies lasiocarpa*	subalpine fir	189	164
ABMA	*Abies magnifica*	California red fir	214	121
ABPR	*Abies procera*	noble fir	70	80
ABSH	*Abies shastensis*	Shasta red fir	31	46
ACMA	*Acer macrophyllum*	bigleaf maple	291	261
AECA	*Aesculus californica*	California buckeye	26	37
ALRU	*Alnus rubra*	red alder	334	358
ARME	*Arbutus menziesii*	Pacific madrone	305	301
CADE	*Calocedrus decurrens*	incense-cedar	573	398
CHNO	*Chamaecyparis nootkatensis*	Alaska yellow-cedar	32	35
CONU	*Cornus nuttallii*	Pacific dogwood	100	72
JUOC	*Juniperus occidentalis*	western juniper	274	356
LAOC	*Larix occidentalis*	western larch	210	212
LIDE	*Lithocarpus densiflorus*	tanoak	197	271
PIAL	*Pinus albicaulis*	whitebark pine	57	34
PICO	*Pinus contorta*	lodgepole pine	475	450
PIEN	*Picea engelmannii*	Engelmann spruce	178	145
PIJE	*Pinus jeffreyi*	Jeffrey pine	389	213
PILA	*Pinus lambertiana*	sugar pine	515	324
PIMO	*Pinus monophylla*	singleleaf pinyon	58	76
PIMONT	*Pinus monticola*	western white pine	255	184
PIPO	*Pinus ponderosa*	ponderosa pine	1133	958
PISA	*Pinus sabiniana*	California foothill pine	48	73
PISI	*Pinus serotina*	pond pine	76	73
POBAT	*Populus balsamifera*	black cottonwood	60	35
PSME	*Pseudotsuga menziesii*	Douglas-fir	1830	1920
QUAG	*Quercus agrifolia*	California live oak	48	31
QUCH	*Quercus chrysolepis*	canyon live oak	407	374
QUDO	*Quercus douglasii*	blue oak	56	97
QUGA	*Quercus garryana*	Oregon white oak	101	106

Table 3 Continued.

Code	Latin name	Common name	Prev-alence (old)	Prev-alence (new)
QUKE	*Quercus kelloggii*	California black oak	463	317
QUWI	*Quercus wislizenii*	interior live oak	84	94
SESE	*Sequoia sempervirens*	redwood	44	76
TABR	*Taxus brevifolia*	Pacific yew	116	81
THPL	*Thuja plicata*	western redcedar	332	321
TSHE	*Tsuga heterophylla*	western hemlock	567	613
TSME	*Tsuga mertensiana*	mountain hemlock	180	161
UMCA	*Umbellularia californica*	California-laurel	97	98

Table 4. Definitions for climate variables derived by Ohmann and Gregory (2002) from Daymet values (Thornton et al. 1997) used in the Principal Components Analysis.

Variable	Definition
ANNPRE	Natural logarithm of mean annual precipitation (mm)
ANNSWRAD	Annual average of the total daily incident shortwave radiative flux (MJ m^{-2} day^{-1})
SMRTMP	Mean monthly summer temperature for June, July, and August (^{o}C)
SMRPRE	Natural logarithm of mean precipitation from May through September (mm)
CVPRE	Coefficient of variation of mean monthly precipitation for wet and dry months (December and July)
ANNGDD	Average number of growing degree days where daily air temperatures exceed 0.0 °C
SMRTP	Moisture stress during the growing season; a ratio of mean summer temperature (SMRTMP) over mean summer precipitation (SMRPRE)
ANNVP	Annual mean of the daily average of partial pressure of water vapor in the air near the surface (Pa)
AUGMAXT	Mean maximum temperature in August (^{o}C)
DECMINT	Mean minimum temperature in December (^{o}C)
DIFTMP	Difference between AUGMAXT and DECMINT (^{o}C)
CONTPRE	Percentage of mean annual precipitation falling June through August

Table 5. PCA loadings corresponding to the Daymet climate variables for each component or eigenvector. We used (and show) the first four eigenvectors or V vectors as climatic predictors where each is scaled to its standard deviation. The PCA was based on a matrix of correlation coefficients among the data. Each respective eigenvalue is shown in the second to last row, and the cumulative variance explained with the addition of each eigenvector is shown in the last row.

Climate Variable	PCA1	PCA2	PCA3	PCA4
ANNGDD	0.9228	0.2589	0.2695	0.0319
ANNPRE	-0.8082	0.5282	-0.0299	0.1217
ANNSWRAD	0.7613	-0.1694	-0.4915	-0.2046
ANNVP	0.2223	0.7232	0.6341	0.1060
AUGMAXT	0.9291	-0.1634	0.1353	0.2793
CONTPRE	-0.1127	-0.7740	0.6006	-0.0894
CVPRE	-0.0056	0.7378	-0.6187	0.2054
DECMINT	0.6329	0.6928	0.3103	-0.0735
DIFTMP	0.3694	-0.8390	-0.1563	0.3644
SMRPRE	-0.9135	-0.0095	0.3162	0.1599
SMRTP	0.9703	0.0903	0.0506	-0.0096
Eigenvalue	5.323	3.260	1.702	0.361
Cumulative % variance	48.39	78.03	93.49	96.78

APPENDIX A: SUPPORTING FIGURES

Figure A1. Relationship of minimum tree diameter versus plot size for the different inventory approaches in the study area. For all inventories, smaller-diameter trees are measured on smaller plots.However, the probability of inclusion of a particular tree size varies among inventories. Legend codes match those used in Tables 1 and 2.

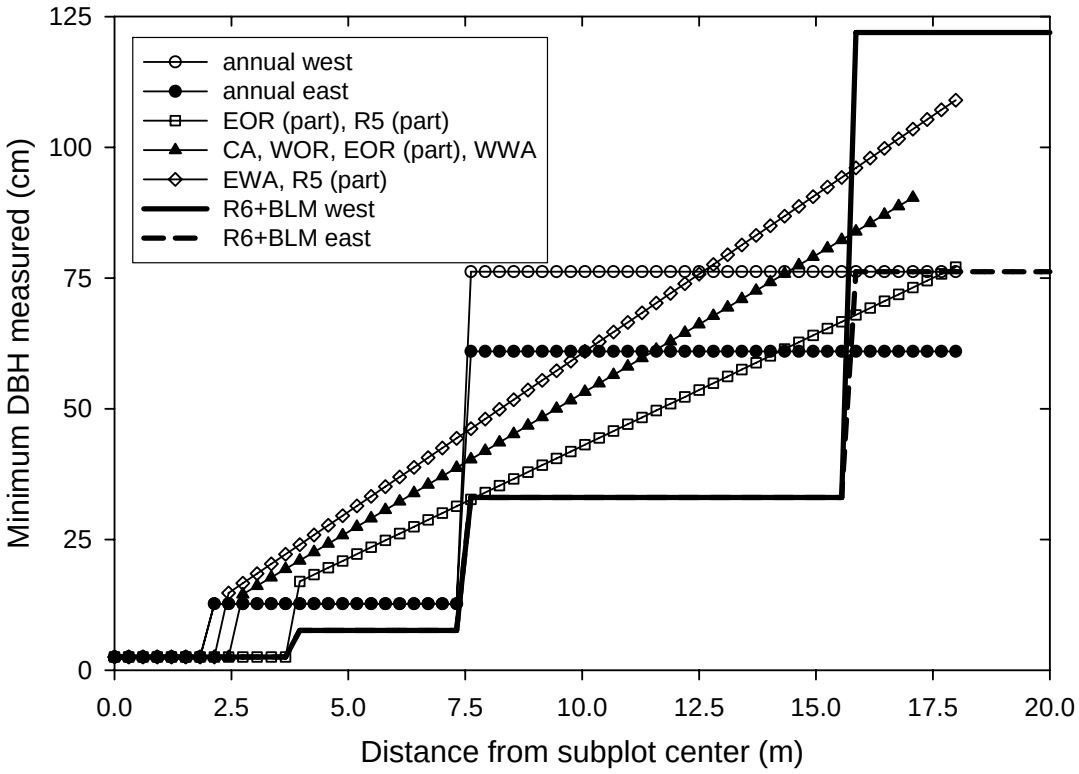


Figure A2. Externally-validated AUC is plotted versus externally-validated $\overline{LogB}$ for comprehensive lists of candidate models generated from NPMR (top and bottom). Values are shown for models generated for three species from the first training sample of the new design, *Arbutus menziesii* (ARME), *Tsuga heterophylla* (TSHE), and *Pinus ponderosa* (PIPO). The numbers of presences varies per species: 301 for *Arbutus menziesii*, 613 for *Tsuga heterophylla*, and 958 for *Pinus ponderosa*. The top axes color code values by species. The color in the bottom axes show the number of predictors or independent variables going into each model.

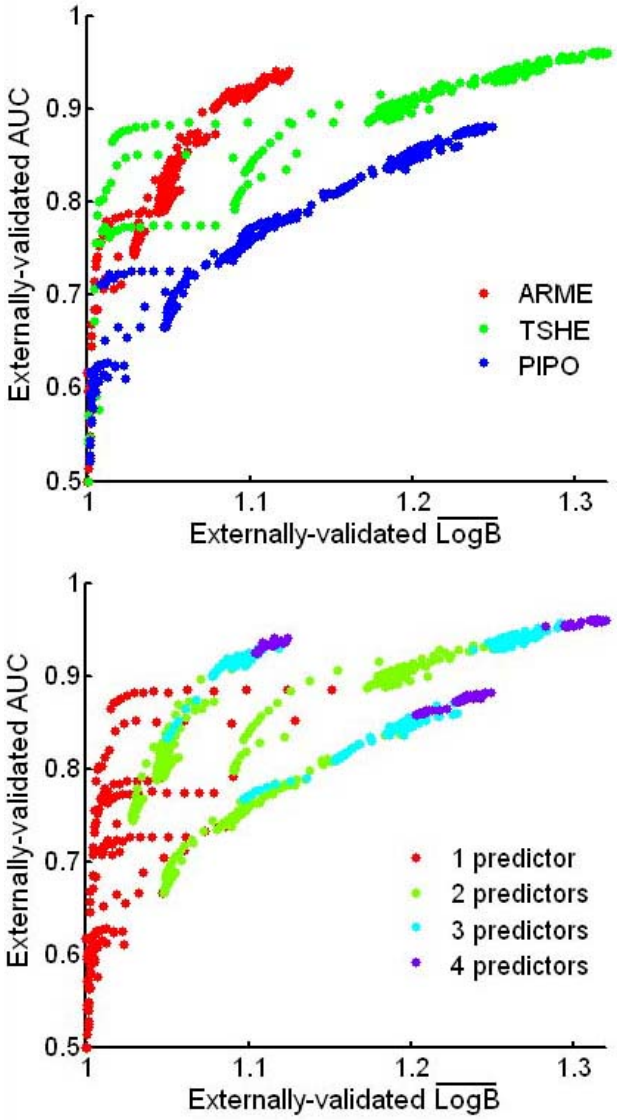


Figure A3. Data were simulated to test the two-sample Kolmogorov-Smirnov statistic (*d*), a measure of effect size, for determination of "climatic bias" among samples of different sizes. 6950 values were generated from normal and bimodal distributions (panel A and B). A pair of random subsamples were taken, one from each distribution, to calculate *d*. This was repeated for 200 replicates across the different sample sizes shown. We plot the means across sample sizes for both comparisons: normal with normal, and bimodal with normal (Figure 5C). The dotted lines represent 95 percent quantiles for the distribution of replicates at each sample size.

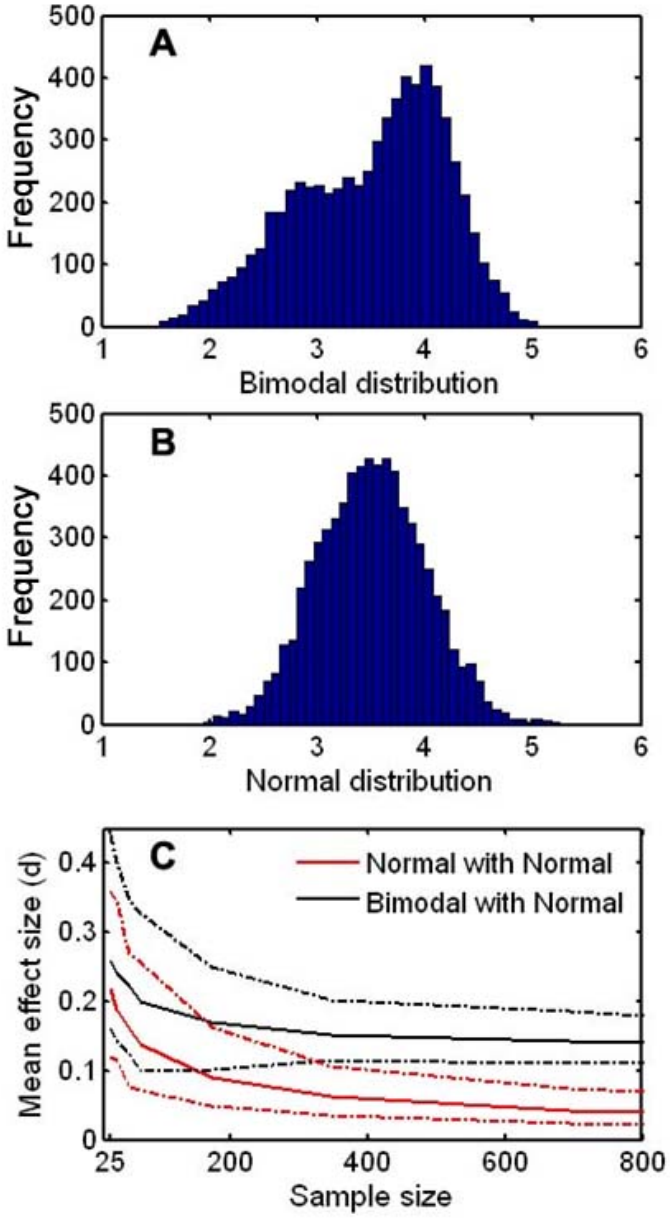


Figure A4. ECDFs (empirical cumulative distribution functions) in the bottom row are generated from simulated distributions shown in the top row at different sample sizes. The figure illustrates the effect of sample size on each ECDF. Small samples yield more jagged ECDFs with a greater likelihood of absolute error among any two compared distributions. The figure also shows how the shape of the ECDF reflects the corresponding shape of the different types of frequency distributions.

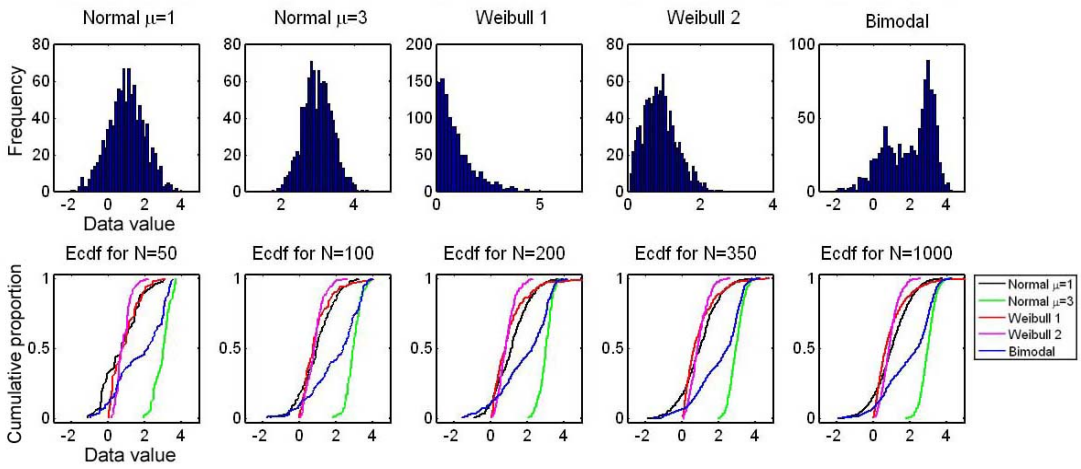
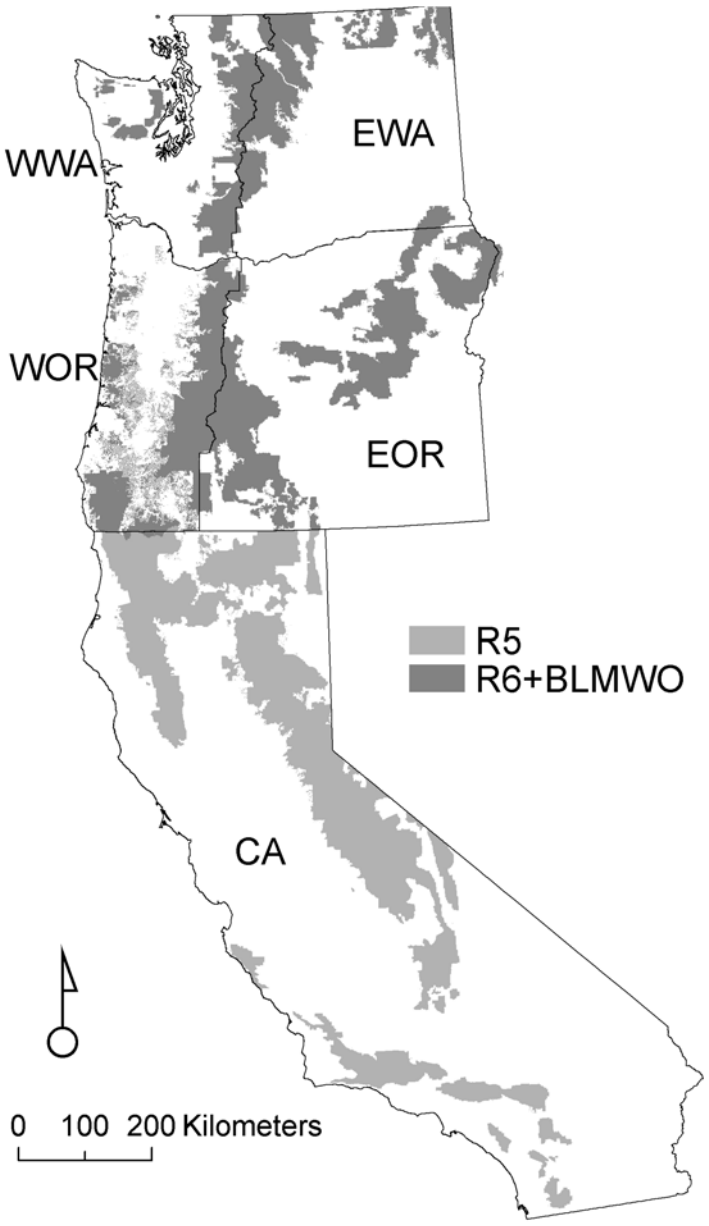


Figure A5. Management regions are shown representing different inventory approaches lumped across the study area for the old inventory. Five distinct geographic regions are demarcated and labeled on the map with boundaries shown by a thin black line (regions EWA, WWA, EOR, WOR, and CA), and two regions are shown by shading (see legend)(R5 and R6+BLMWO). (B2).



APPENDIX B: SUPPORTING METHODS

Examination of climatic bias with ECDFS and QQ-plots. Our characterization of climatic bias from the data requires explanation. We considered using a measure of effect size, the Kolmogorov-Smirnov Two-Sample Test statistic (*d*) (Massey 1951), to compare climatic bias across different sample sizes (or species with different prevalences). This statistic assumes no particular form between the compared distributions, it measures the maximum absolute difference among empirical cumulative distribution functions, and it accommodates differences in both shape and central tendency. However, we checked the immunity of this statistic, *d*, to sample size, and we discovered *d* to depend on sample size using simulated data (see Appendix A; Fig. A3). The shape of the dependence varies with the distributions being tested (Fig. A3). As sampling differences can be reflected not only in the mean but in the shape of a frequency distribution, we instead simply plotted the empirical cumulative distribution functions (ECDFs) from the old and new data sets to visualize the climatic bias per species (Chambers et al. 1983). For each observed value in a distribution, the empirical ECDF plots the fraction of points that are less than the observed value. Numerous ECDFs can be easily condensed and shown in tandem, and they represent the mean, standard deviation, and standardized third and fourth moments all in one figure (Wilk and Gnanadesikan 1968). We also used quantile-quantile plots (QQ-plots) of two distributions to further investigate evidence for climatic bias with four species. QQ-plots show empirical quantiles from two samples plotted against each other to determine if they come from the same distribution (Chambers et al. 1983). QQ-plots are a powerful approach to zoom in and compare shape of distributions underlying two samples of data.

Chapter 4. Do Climate Extremes Explain More Variability than Means in Tree Species' Occurrence by Life Stage?

Heather E. Lintz, Andrew N. Gray, Bruce McCune

ABSTRACT

We pit climate means against climate extremes to determine species' probability of occurrence by life stage for tree species across the Pacific coastal United States. We use climate means and extremes for variables that have direct bearing on plant fitness, growth, and mortality. Variables include growing season length, seasonal vapor pressure deficit, seasonal minimum temperature, degree of freeze-thaw events occurring in the spring and fall, and seasonality of precipitation. These are derived from PRISM climate data at a monthly time step across the period from 1900 to 2007. We derive climate extremes by first testing for differences among time intervals in climate with Multivariate Response Permutation Procedure or MRPP. We calculate average measures of climate deviation per time interval defined by combinations of climatic oscillations, which we call oscillation schemes. We build niche models for presence/absence data of seedlings and adults across 22 tree species using data from USDA Forest Inventory and Analysis (FIA). To build a model for a species and climate variable, we supply the long-term climate mean and related extremes to the model building process using Non-Parametric Multiplicative Regression (NPMR). This process is repeated for 22 species, two life stages, and seven climate variables. Results indicate that 27% of the models had a climate extreme explain more variability than a climate mean. Extremes associated with freeze-thaw events, seasonality of precipitation, and winter minimum temperature were most frequently represented. This is suggestive evidence that extremes were less often chosen for models of seedling niches compared to adults ($p=0.054$). The reason for this is counterintuitive and unclear. Perhaps the adult niche has a longer "memory" to climatic events and thus a higher likelihood of being shaped by climate extremes.

INTRODUCTION

Patterns of species' distributions across the landscape may be the most pronounced but least understood features of life on earth. Such patterns have long compelled scientific study and continue to do so especially with the pressing topic of climate change. The concept of the "species' niche" underpins much of the scientific

inquiry related to species' distributions. The niche concept was founded in the early twentieth century and periodically strengthened among various authors (e.g. Grinnell 1917, Gause 1936, Hutchinson 1957, Leibold 1995). The intersection of these authors defines a niche as the quantitative description of conditions that allow a species to occur. The niche can be expressed as either 'fundamental' or 'realized.' The fundamental niche describes the full range of environmental conditons under which an organism can exist (e.g. under controlled conditions in the laboratory) while the realized niche results from natural constraints such as interactions with other organisms in a community (e.g. superior competitors). The status quo for quantitative descriptors of the niche is the use of mean conditions such as mean annual temperature.

The relationship between species' occurrence and mean condition is the basis for statistical models, and a recent surge of forecasts rely on species' occurrence as a function of climatic means (e.g. Peterson et al. 2002, Shafer et al. 2001, Rehfeldt et al. 2006). However, niche models using means of conditions can be oversimplified (Jackson et al. 2009, Chase and Leibold 2001, Colwell et al. 2009). Also, the success of means of climatic conditions in niche models often relies on covariance structure among climatic variables, which will change with broadly changing patterns of atmospheric circulation (Jackson et al. 2009). Further, climatic extremes will likely increase and have profound effects on populations that translate to niche structure (Gutschick and Bassirad 2003, Boyce et al. 2006, Gomulkiewicz and Holt 1995, Allen and Breshears 1998, Stenseth et al. 2002). Given the imperative for understanding biological effects of climate extremes with climate change (e.g. Gutschick and Bassirad 2003, Stenseth et al. 2002, Parmesan et al. 2000), the broad-scale study of species' niches with respect to extremes in climatic conditions is clearly warranted.

A recent blend of paleo-ecological and paleo-climatic work supports that modes of historical climatic variability synchronize with pulses and retreats of species' occurrence (Swetnam and Betancourt 1998, Swetnam et al. 1999, Gray et al. 2003, Jackson et al. 2009). Pulses of establishment are thought to require unique climatic circumstances because environmental responses of species often depend on life stage

(e.g. Grubb 1977, Ibanez et al. 2009, Chaubert-Pereira et al. 2009, Jackson et al. 2009). In the case of tree species, the juvenile fundamental niche is thought to be narrower compared to that of adults (e.g. with respect to a mean climatic parameter) and thus more susceptible to climate extremes associated with climate oscillations (Jackson et al. 2009). These ideas are fundamental to processes that contribute to species' occurrence. However, experimental tests of these ideas are not possible at the landscape scale. Instead, statistical tests for consistency in relationships through time play an important role in understanding macro-ecological process (Kerr et al. 2007). Here we perform statistical tests to complement ideas emerging from historical patterns. We ask if climate oscillations, as expressed in extremes, can explain substantial variability in species' occurrence. We also test the null hypothesis that the frequency of extremes explaining more variability than means does not depend on life stage. The alternative to this hypothesis addresses the possibility that seedlings are more susceptible to extremes than adults.

METHODS

Overview. Our method involves four general steps. Each step is described in each of the four following sections. Here, we provide an overview. *First*, we obtain our response variables from the Forest Inventory and Aanlysis (FIA) database as presence/absence data by tree species and by life stage (see section entitled Study Area and Tree Species' Data). *Second*, we derive climate predictors associated with the FIA plot locations that provide the presence/absence data (described under Climate Analysis). Climate predictors include climate means and extremes. We turn to standard climatic oscillations as a way to group and summarize the climate extremes. Before deriving the extremes, we first test for a winning 'oscillation scheme' that best segregates the values of climate anomalies for each of seven climate variables (see the section below entitled Climate Analysis). We then use the groups of years defined by each winning oscillation scheme to summarize anomalies by taking the mean for the group. We define extremes as the mean anomalies during particular oscillatory phases, i.e. segments of time. We

perform this process separately for the seven climate variables. *Third*, we build statistical models by species and life stage to determine if more variability is explained by climate extremes or climate means (described under Niche Modeling). We regress species' presence/absence against climate mean and related extremes for each climate variable, one at a time. We repeat this for each of two life stages and for each species. *Fourth*, we use the results from the niche modeling to test the hypothesis that the frequency of extremes explaining more variability than means does not depend on life stage (described under Hypothesis Testing). Also, to give account to the differences in numbers of predictors supplied for each model, (e.g. one mean and *x* number of extremes where *x* varies per climate variable and life stage), we perform a second test. We determine the probability of a result as extreme or more extreme as the numbers of means chosen over extremes per climate variable and life stage (assuming all variable have an equal chance of being chosen).

Study Area and Tree Species' Data. The study area is the U.S. Pacific coastal states of California, Oregon, and Washington. Diverse climates and biomes occur bearing the largest range of mean annual rainfall in the conterminous United States. The region has the highest coniferous tree diversity and endemism second to Mexico. Evidence of increased mortality and migration for tree species exists in the region for the past decades (van Mantgem et al. 2009, Kelly and Goulden 2008).

Plots containing at least 50% forest are selected from FIA data. This criterion helps to avoid species' absences caused by factors other than climate, such as agriculture. Presence/absence data from seedlings and adult trees are segregated. The distinction of tree versus seedling is made in the field by size. Seedlings are defined as having a diameter of less than 2.5 cm at root collar and length greater than either 12.7 cm for conifers or 30.5 cm for hardwoods. Thus, the "seedlings" are beyond the initial stages of heavy stochastic mortality occurring typically in the first year or two. Adult trees were defined as having a diameter at breast height or root collar of greater than 12.7 cm. Seedlings were sampled from the total area of 54 m^2 per plot; whereas, adult trees are sampled from an area totaling 672.5 m^2 per plot. Seedling prevalence (or number of

presences for a data set) is generally fewer compared to adult trees due to a smaller probability of detection. Presence/absence data from across the years 2000 through 2007 are pooled. Only species with more than 200 presences of seedlings are retained among the seedling and adult data sets (Table 1). To examine preliminary evidence of niche differentiation between life stages, we count FIA plot locations containing just seedlings, just adults, and both seedlings and adults for our study species. We visually compared these counts. We also created data sets to build niche models. The seedlings and adult data pools were randomly sampled without replacement to serve in the model building (N=2500). Figure 1 shows the locations for the training data set (N=2500).

Climate analysis. One grand challenge in studying niches and climate extremes is how to characterize the extremes. Below we describe methods to discover and use discrete time intervals for the purpose of summarizing extremes from the perspective of a tree. We select climate drivers of tree species' occurrence based on previous works (see Table 2 for definitions and justification). Although the long-term means of some variables are strongly correlated (for example, growing season length and the minimum winter temperature), their deviations or extremes are much less so as we further demonstrate. We test whether discrete groupings of years based on large-scale oscillations (such as the Pacific Decadal Oscillation) are suited to characterize extremes in the climate drivers. To this end, we classify years within the time interval from 1900 to 2007 according to indices of climate oscillations. We explore all possible combinations of the Pacific Decadal Oscillation (PDO) (Mantua et al. 1997), the Atlantic Multi-decadal Oscillation (AMO)(Enfield et al. 2001, Schlesinger and Ramankutty 1994), and the El Niño Southern Oscillation (ENSO)(Trenberth 1984, Trenberth and Stepaniak 2001, Trenberth 1997)(Fig. 2, Table 3). The PDO, AMO, and ENSO influence drought, temperature, and precipitation in the western United States (McCabe et al. 2008, Enfield et al. 2001, Sutton and Hodson 2005, Benson et al. 2003, Ropelewski and Halpert 1986, Cayan et al. 1998, Zhang and Delworth 2007). The different oscillations interact or modulate one another (Enfield et al. 2001, McCabe and Dettinger 1999, Zhang and Delworth 2007, McCabe et al. 2008).

We test three indices of ENSO as each represents different phenomena (Fig. 3, Table 3) (Trenberth and Stepaniak 2001; Trenberth and Hoar 1996). We test the Southern Oscillation Index (SOI) (Trenberth 1984), the Trans-Niño Index (TNI) (Trenberth and Stepaniak 2001), and the Niño 3.4 Index (N3.4) (Trenberth and Stepaniak 2001, Trenberth 1997). We aimed to discern years with a predominance of months above or below a threshold. Consequently, we used an annual mean threshold to classify ENSO events. We test two threshold levels for each ENSO index since the threshold levels are subject to debate (e.g. Trenberth 1997) (Table 3). We also include a time interval called "R1" as a grouping variable discerning the period from 1976 to 2007 (Fig. 2), which is characterized by more frequent and longer El Niños (Trenberth and Hoar 1996), rapid temperature rise in many areas of the world (IPCC 2001), and a switch in the behavior of the TNI (Trenberth and Stepaniak 2001). After classifying years across the past century, we create a list of different classification schemes to test as grouping variables for climate anomalies (Table 4, Fig. 2). The schemes comprise oscillations alone and in combination. We compare the efficacy in group separation of anomalies among schemes separately for each climate variable. PDO and AMO and their combinations are treated either as unique groups through time (hence the label 'gr') or simply as phases recurring in the same category (e.g. PDO extremes at any time, not just a particular sets of contiguous years)(Table 4, Fig. 2). The unique grouping expresses climate oscillations as a unique events rather than a recurrent fixed mode. However, unique grouping is only possible for the lower frequency oscillations, otherwise, our purpose of dimensionality reduction would be defeated.

We randomly sample 500 locations within Oregon, California, and Washington. We extract time series for each location using monthly precipitation, temperature, and dew point temperature from interpolated climate grids at a resolution of 4 km (PRISM; Daly et al. 2002). The time series date from 1900 to 2007. We derive our set of climate variables from the time series for each location (Table 2). We calculate annual 'deviations' for each variable by subtracting the long-term mean of 108 years from annual-scale values.

We determine the strongest grouping variable across time series using Multi-Response Permutation Procedures (MRPP) (Mielke and Berry 2001) (PC-ORD version 5.0; McCune and Mefford 2006). MRPP is a non-parametric substitute for Multivariate Analysis of Variance. It tests the null hypothesis of no difference between two or more groups of entities. MRPP also provides an effect size or the degree of separation among groups, the *A*-value. The *A*-value theoretically has no lower bound and has an upper bound at 1 meaning that all values are identical within groups and different among groups. An *A*-value of zero is the random expectation; 0.1 represents a strong degree of separation for ecological data. Euclidean distance is used as the distance measure.
We summarized deviations using the best grouping scheme per climate variable using the lowest *p*-value and the largest *A*-value among all MRPP tests for a climate variable. The average local deviation is then calculated within each group of years per oscillation scheme and climate variable (see Fig. 4). The calculations are repeated for each FIA plot location. We use Principal Components Analysis (PCA) of the climate data among our plot locations (N=2500) to determine the amount of covariance shared among three different data sets: the long-term means for the seven climate variables, all the extremes across the seven climate variables, and the pooled means and extremes (PC-ORD version 5.2; McCune and Mefford 2006). The PCA used a matrix of correlation coefficients among the data.

Niche modeling. We used Non-parametric Regression (NPMR) to build niche models (Hyperniche 2.07; McCune 2006, McCune and Mefford 2008). This is also known as Non-parametric Regression (NPR) when regression involves only single variables, which is mostly the case here. NPMR is a kernel regression technique and outperforms other popular niche modeling approaches (Lintz et al. 2010, Yost 2008). We use default settings with a local mean and Gaussian kernel. We build eight models for each life stage for each species, one model for each set of climate variables that were derived from a single raw climate variable (Table 2). The derived variables include means and extremes, as described above. The number of available predictors for each model varies by life stage and climate variable. The number of groups for the winning

oscillation scheme (plus one, the mean) determines the number of predictors per analysis. For adult presence/absence, we include all summaries of extremes dating back to 1900. For seedling presence/absence, we select oscillations groups dating back 30 years from the last year of sample (2007). Our goal is to represent extremes experienced across the variation of ages corresponding to that life stage. The scope of available climate grids determined our temporal bound of 108 years for the adults.

NPMR selects the top single-variable models for each analysis. Model selection uses an iterative optimization that chooses the best log likelihood ratio based on cross-validation. However, the log likelihood ratio is sensitive to number of presences in the data, and thus, is not ideal for comparing fit across species with differing numbers of presences. Instead, we use the Area under the Receiver Operator Characteristics (ROC) curve or AUC for model evaluation or measurement of fit (Hanley and McNeil 1982). An AUC of 0.5 represents a model fit no better achieved by chance alone. The maximum value of the AUC is 1.

Hypothesis testing. We compile a contingency table to give the numbers of species with either the long-term mean or extreme chosen by life stage. Fischer's exact test evaluates the null hypothesis that the selection of the long-term mean is independent of life stage for each two-by-two table.

Since the numbers of extremes given to each analysis differed by climate variable and life stage, we perform another test. We calculate the probability of a long-term mean being chosen as frequently as or more frequently than the observed value for each life stage and climate variable. If means are chosen more frequently across species than we would expect by chance given the null model, then the *p*-value will be small. The null model assumes all variables have an equal chance of being chosen. Think of each model constructed as a Bernoulli trial where the two outcomes are either the long term mean is chosen as the best predictor or not, in which case a climate extreme was selected as the best predictor. Each species represented a trial for a climate variable. Because the number of extremes represented as possible predictors depend on the climate variable, the probability of the two outcomes varies for each model. We calculate the likelihood of the

result for each climate variable and life stage using the binomial probability mass function:

$$\Pr(x = k) = \binom{n}{k} p^k q^{n-k}$$

$$\text{where } \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

where p is the probability of success, q is the probability of failure, n is the number of trials (or number of species participating in the test), and k is the number of successes (or number of species with long-term means chosen for a given climate variable). To integrate probabilities that are more extreme than k, we summed probabilities across k to n-k. This test informs us whether observed numbers of means (or extremes) chosen could be due to statistical artifact if the null model holds (where all variables have an equally likely chance of being chosen). Although this null model likely does not hold, it is not impossible and still must be given account.

RESULTS

Climate analysis. Climate anomalies for each climate variable were successfully grouped by oscillation schemes as shown by *A*-values and small *p*-values from MRPP (Table 5, Fig. 3). The strongest oscillation scheme differed across climate variables, and the strength of separation among groups of years also varied. Growing season length and minimum summer temperature were grouped most distinctly by the same oscillation scheme (PAgr). This scheme included PDO and AMO phases considered as unique groups through time. R1 played a role in grouping anomalies for five of the seven climate variables. The strongest cohesion among groups of years was for winter and summer minimum temperatures followed by summer vapor pressure deficit and growing season length. Five of the seven climate variables showed the most cohesive groups of years when grouped in part by ENSO indices. Each of the ENSO indices tested was involved

with a winning oscillation scheme with SOI being most frequent. Two of the five winning oscillation schemes with ENSO indices relied on strong ENSO events (with a cut-off of +/- 0.75^{o}C for TNI and SOI), while the others had a more moderate cut-off (+/- 0.4^{o}C for N.34 and SOI). All oscillation schemes involved more than one climate oscillation except for one, which relied on SOI (as modulated by R1) for thaw-freeze events in the spring.

Oscillation schemes showed geographic cohesion for each climate variable to varying degrees, as illustrated by differences among groups of years across all FIA plots (Fig. 4). Some oscillation schemes with relatively low *A*-values from the MRPP (which was a test of segregation in time) still showed marked cohesion across space for the given years (e.g. spring thaw-freeze). The directional effect of El Niño and La Niña events can be reversed depending on decadal-scale phases of the climate system (e.g. see fall freeze-thaw, spring thaw-freeze, and winter minimum temperatures for the best examples). The past three decades (denoted as R+ in Fig. 4) are linked to an increase in directional deviation and variation for some oscillation phases and climate variables. For example, thaw-freeze events in the spring are more pronounced than for R-; however, this effect is only seen during years not affected by ENSO events (Table 5, Fig. 4). Another example is summer vapor pressure deficit. The strongest positive deviations (or greater vapor pressure deficits) are seen for recent years where PDO and AMO are positive and not affected by TNI events. Also, variation across space in summer vapor pressure deficit is the highest among all oscillatory phases for recent years where PDO and AMO are positive. Increased variability in anomalies can also be seen for other variables where R1 is positive (e.g. seasonality of precipitation, spring thaw-freeze, and winter minimum temperature).

Species' analysis. Seedlings and adult trees only partially overlap in their plot occurrences (Fig. 5). To examine whether it is necessary to filter out poorly fitting models to ascertain our conclusions, we first report results for the binomial probability test. The probability that the observed numbers of winning models based on means was due to chance alone was very small for all models and climate variables (Table 8). Also,

the converse was true for the probability that extremes would be selected over means. There is a high probability that the observed number of extremes would be selected by random chance alone for every climate variable assuming all variables have equal chance to be chosen (e.g. p-values would be virtually identical to 1-p for ps reported in Table 8). Consequently, we filter out poorly-performing models in reporting our results to safeguard against the possibility that extremes are selected by random chance alone. Although the AUC itself is assurance where an AUC of 0.5 or lower is model fit no better achieved by chance, we opt for a conservative cut-off in the AUC (0.75) and examine only the best-performing models. We assume that the null model where all predictors have an equal chance of being chosen does not apply especially for models with fits greater than or equal to an AUC of 0.75.

Extremes explained more variability than means across life stages for 27% of a model pool where poorly fitting models were filtered out (Table 7). We rejected the null hypothesis that the proportional selection of the long-term climatic mean is independent of life stage for a model pool where poor fitting models were filtered out (p=0.054, from Fischer's Exact Test). (This result was corroborated by the same test across all models, p=0.004.) Models for large trees more often showed extremes explaining more variability than means. However, long-term means were still the most frequently selected for both life stages.

The number of species represented among best-fitting models was greatest across life stages for seasonality of precipitation, summer vapor pressure deficit, and spring thaw-freeze (Fig. 6). When median AUCs in model fits for top-performing models were visually compared among life stages, model fits for seedlings outperformed adults for growing season length, winter minimum temperature, and summer vapor pressure deficit. Whereas, adult models outperformed seedlings models with respect to summer minimum temperature, spring thaw-freeze, fall freeze-thaw, and seasonality of precipitation. Of the extremes, fall and spring freeze-thaw (for both life stages), seasonality of precipitation (for large trees), and winter minimum temperature (for large tree) were represented in

greatest proportion for the best-performing models showing more variability explained by extremes (Fig. 6).

DISCUSSION

A climate extreme defined according to climate oscillations can explain more variability than a climate mean for tree species in the Pacific coastal United States; however, this circumstance occurred for a minority of the best-performing models with good fits across species and life stages (27%). This result corroborates decades of studies demonstrating the predictive power of climate means while offering an important caveat. Although climate means carry substantial predictive capacity, extremes can carry more predictive power in approximately one-quarter of circumstances for the study species, climate variables, and time period we examine.

Interestingly, seedling niches did not show greater sensitivity to extremes compared to adults; this was expressed by means winning over extremes more often for niches of seedlings compared to adults. The reason for this is unclear. We discuss various possible explanations.

Studies have shown that tree seedlings are more vulnerable than adults to climate fluctuations. For example, some seedlings can have greater moisture requirements and sensitivity in growth rate to climate fluctuations (Ibanez et al. 2009, Tanaka et al. 2008, Chaubert-Periera 2009). To be more sensitive to climate extremes, the fundamental climate niche for seedlings with respect to the climate mean should be narrower compared to an adult niche (Jackson et al. 2009). Realized seedling niches evident from our study area differed from adults slightly in shape (Fig. 7), and similar patterns are corroborated in other temperate regions (e.g. Lenoir et al. 2009). However, realized seedling niches found across tree species were not notably narrower than adult niches across models (e.g. to the degree conceptualized foe the fundamental niche by Jackson et al. 2009). Despite the fact that the traits of the seedling life stage and aspects of its realized niche with respect to a climate mean can differ from adults, this difference did not translate to a stronger relationship to extremes for our study region and time period.

Since seedlings are smaller than adults, microhabitat can offer shelter that may subsidize some physiological draw-backs. The distribution of the saguaro cactus exemplifies this with respect to low temperature tolerance. Winter freezing has been proposed as a limit to the distribution of the saguaro (Shreve 1911). However, winter freezing was shown to differentially affect (and kill) adults compared to juveniles because juveniles were sheltered by rocks and nurse plants (Niering et al. 1963). It is reasonable to suspect that similar processes associated with microhabitat protection may afford tree seedling niches greater immunity to some extremes. However, species' niches at the landscape scale of study are not amenable to *in situ* experimentation, and testing such a hypothesis is still a major research challenge.

To better understand the situation where adult niches related to extremes and seedling niches did not, we visually explored modeled results where the long-term mean for a seedling niche was selected while an extreme was selected for the adult niche (Fig. 8). This was the case for five species where model fits for both seedling and adults were greater than or equal to an AUC of 0.75 for the same climate variable. After first force-fitting the adult data to the long-term mean to compare with seedling niches, and noting very little differences (Fig. 7), we turned to the best two-predictor models for the adult trees.

We show response surfaces of probability of occurrence for adult trees versus the long-term mean and the extreme (Fig. 8). The extremes modify relationships with the long-term means. In every case, an increase in probability of absence (that would otherwise not be detectable from a single-variable model) occurs with respect to the long-term mean at certain levels of the extreme. For example, absences occurred with negative deviations in the SI index where precipitation became less seasonal than the norm for three species, *Abies amabalis*, *Thuja plicata*, and *Tsuga heterophylla*. Also, *Calocedrus decurrens* occurs where negative deviations in spring thaw-freeze are strongest, or freeze-thaw events are least severe (with respect to the x-axis), and absences tend to occur where deviations are weakest. Finally, *Lithocarpus densiflorus*, which is most often found in areas with long growing seasons, is absent for smallest negative deviations from growing

season length. Occurrence is associated with both a long growing season and the strongest negative deviations in growing season length for P-A-1622, which also happens to be the strongest negative deviations for the century (see Fig. 4).

The modeling method we use, NPMR, can capture interactions that occur at any temporal and spatial scale in the data. So, if a climate extreme influenced probability of occurrence for a certain climate interval and portion of a species' range, it can be captured and visualized. However, the strongest models in species occurrence given climate likely involve more than one climate variable due to interactions. The inclusion of different climate variables together as means and extremes in the process of building niche models is an intriguing topic of future research.

The most compelling explanation for seedlings showing more immunity to climate extremes compared to adults is that the adult niche has a longer "memory" to climatic events. Each extreme has a unique temporal and spatial expression. Vulnerability to extremes depends both on the degree of the extreme along with its spatial and temporal expression relative to a tree's distribution and phenology. Thus, a longer time period of life allows for greater chance that a climate extreme could influence a species' niche and spatial distribution across the landscape.

Climate considerations. Our freeze-thaw indices contribute to a growing body of work that examines the ecological influence of freeze-thaw events at different temporal scales (e.g. occurring across days to weeks to months)(e.g. Neilson and Wullstein 1983, Jentsch 2007, Gu et al. 2008). We developed indices to measure transitions of anomalous freeze to anomalous thaw (in the fall) or vice versa (in the spring) at a monthly time step. We demonstrate that the indices are ecologically meaningful. We provide evidence for the influence of freeze-thaw events across tree species' niches and ranges in the Pacific coastal United States. For example, extremes in either spring thaw-freeze and fall freeze-thaw explained more variability than means for twelve species (where only best-fitting models were considered).

CONCLUSIONS

Tree species' niches not only appear to respond to climate extremes defined by climate oscillations, but extremes can be a dominant explanatory factor for some species. Additionally, seedlings niches may not be as sensitive to climate extremes as previously thought. A possible explanation is that the adult niche has a greater chance of being affected by climate extremes because of a longer life span. Our results bear on the interpretation of projections of species' occurrence given climate change. Currently, species' projections are made using climate model forecasts. However, climate science does not yet understand decadal to century scale climate variation; nor can climate models replicate decadal to century scale climate variation (Warren Washington, personal communication). Hence, we face a grand challenge in forecasting tree species' occurrence. If some tree species are affected by extremes modulated by decadal-scale climate variation, and we are shy in understanding how decadal-scale climate variation will unfold, then we should make and use tree species' predictions in future climates with caution.

ACKNOWLEDGEMENTS

We thank the U.S.D.A. Forest Service for funding this work.

FIGURES

Figure 1. Map of Forest Inventory and Analysis plot locations used in model building from across Oregon, California, and Washington (N=2500).

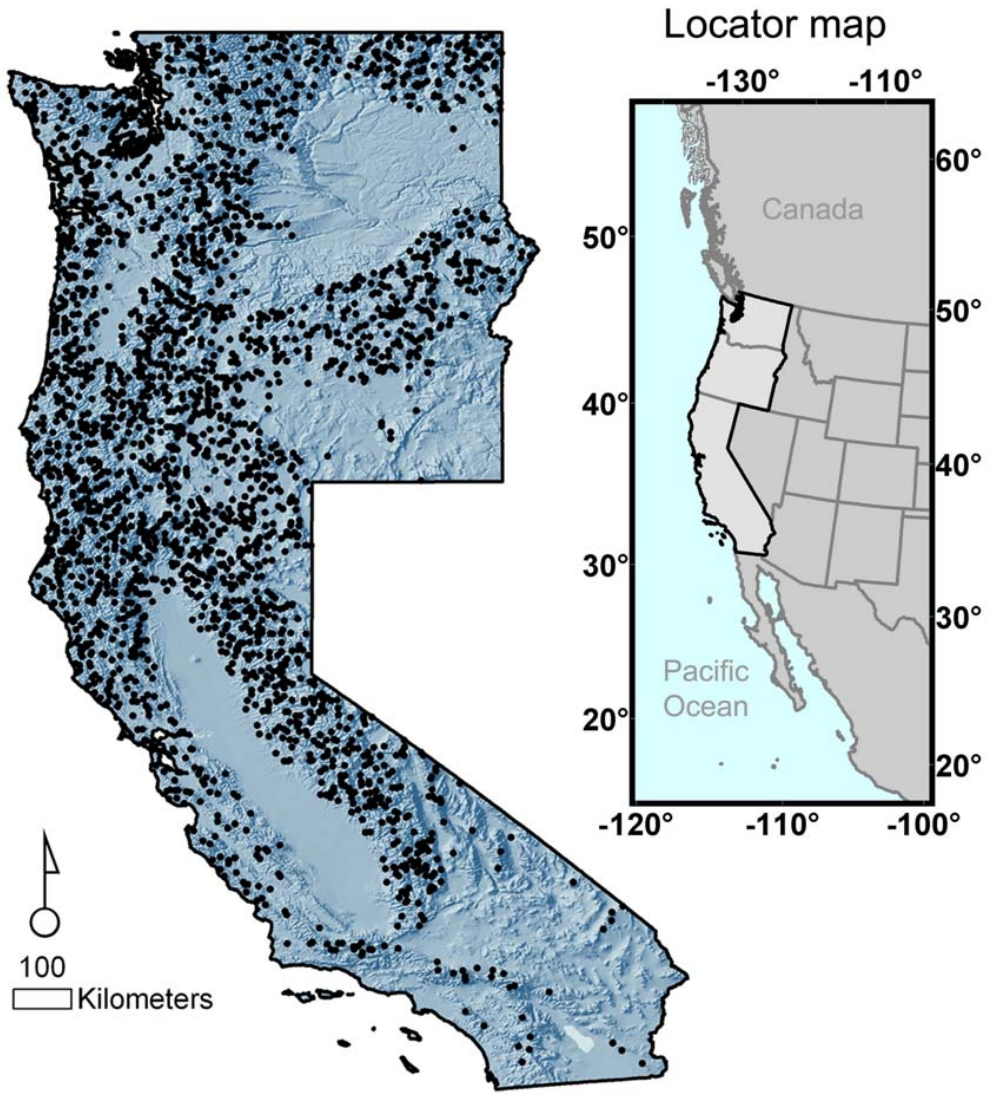


Figure 2. Different grouping variables representing decadal scale oscillation schemes are shown. Different schemes are developed through group intersection (e.g. PDOAMO from PDO and AMO). Each decadal-scale scheme was treated either as unique groups through time (hence the label 'gr,' right) or simply as different phases having the same category (e.g. PDO and AMO; top). Each color represents a unique category or group within a grouping variable. Colors do not correspond across groups. Time is represented on the x-axis.

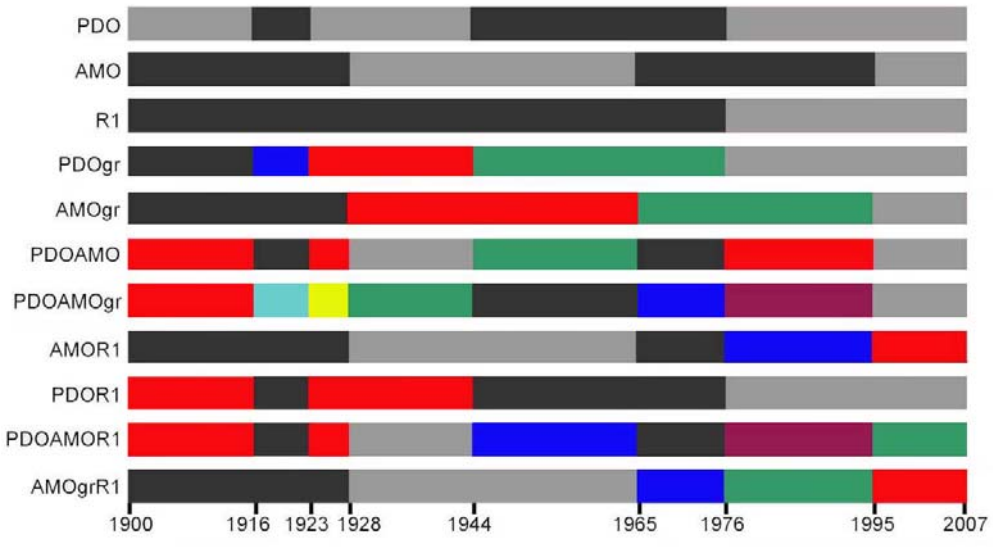


Figure 3. Winning oscillation schemes are shown (corresponding to Table 5). Colors schemes for each bar are not shared among bars. Color meanings are shown above. P=PDO, A=AMO, R=R1 (see Tables 3 and 4 for definitions).

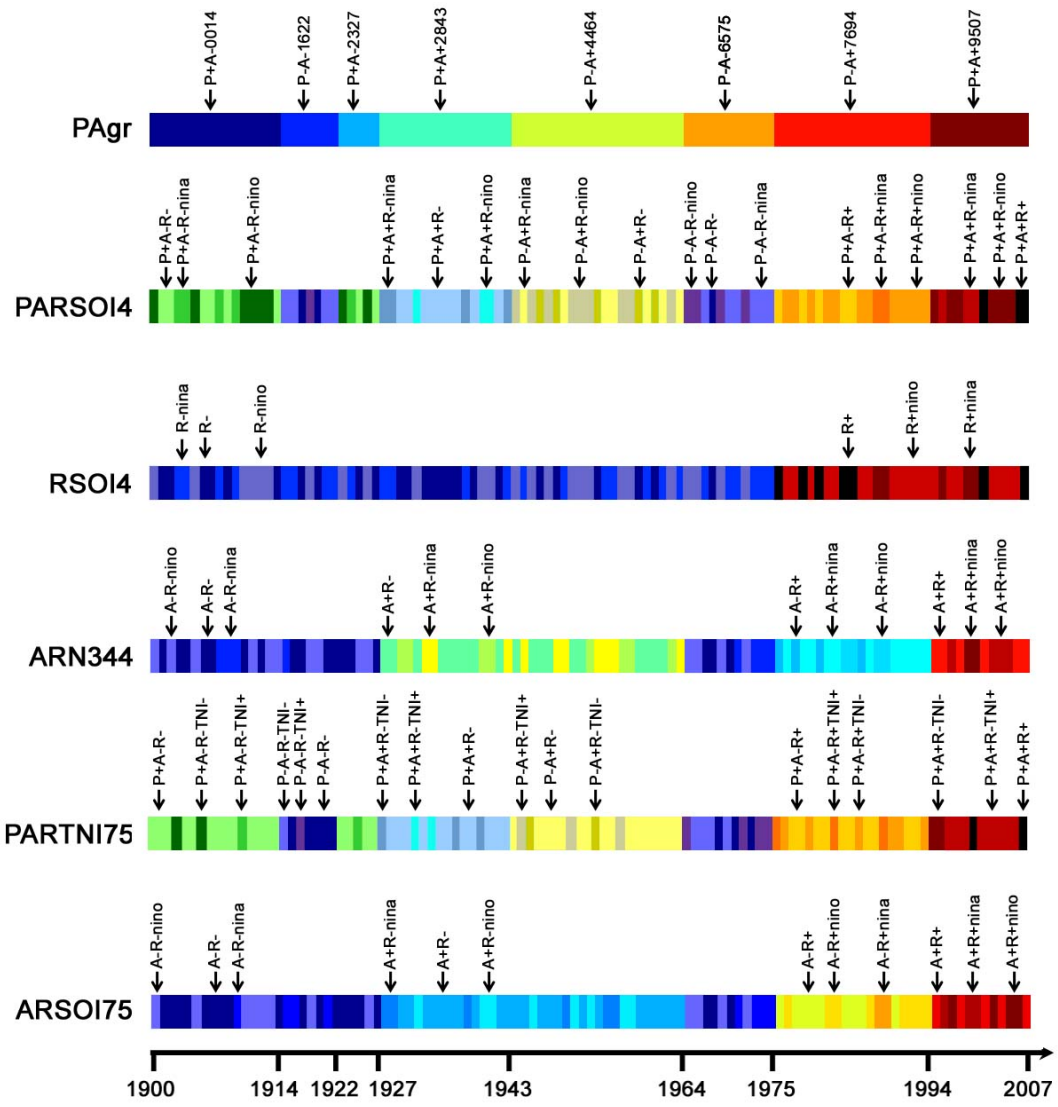


Figure 4. Quantiles for anomalies averaged within phases (or groups) within winning oscillation schemes. The winning scheme for each climate variable (*x*-axis) is identified at the top of each graph. Each code on left side indicates a particular time period or recurrent phase within the winning oscillation scheme. Quantiles are shown for FIA plot locations contributing to model building (N=2500). Results for all phases within each oscillation scheme unique to a climate variable are shown (see Table 5 and Fig. 3). The open circle represents the median, dark circles are first and third quartiles, and hatch marks define the bounds of 2.5 and 97.5%.

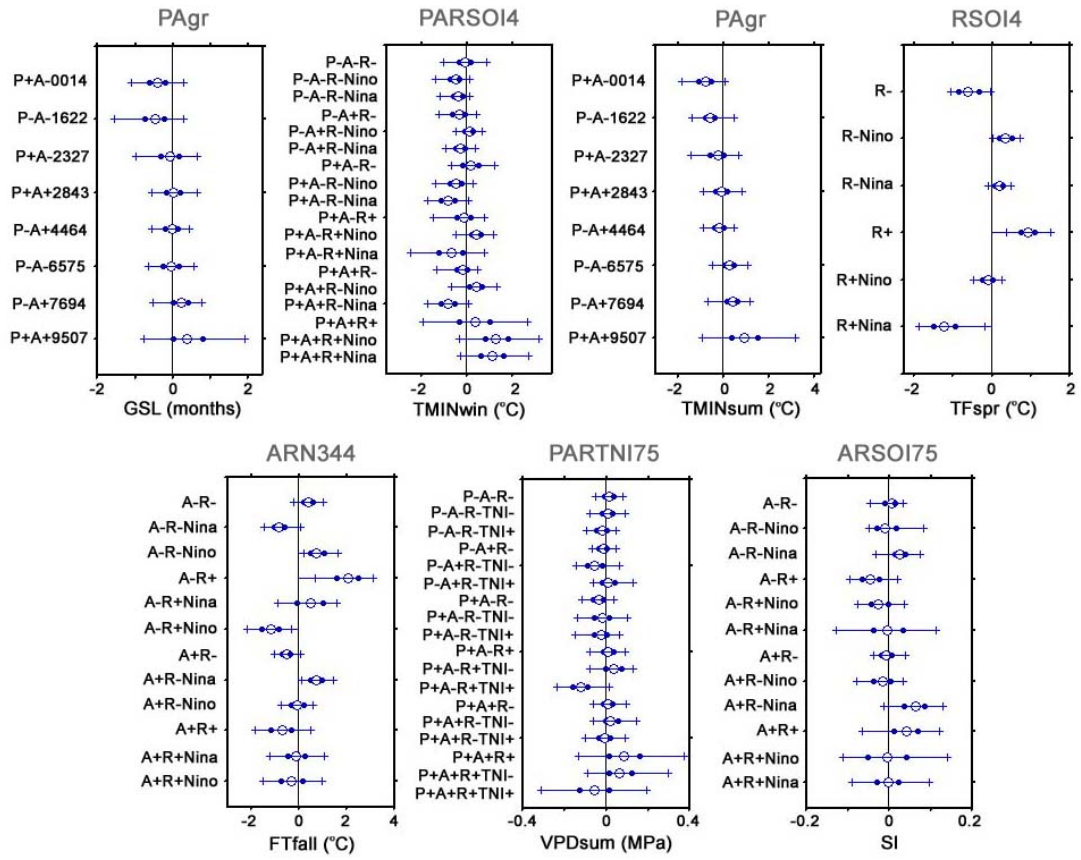


Figure 5. Plot counts across Oregon, California, and Washington by life stage and species. Columns separate plot groups (labeled above). PSME and PIPO are given different axes due to their dominance. Counts span the time period from 2000 to 2007.

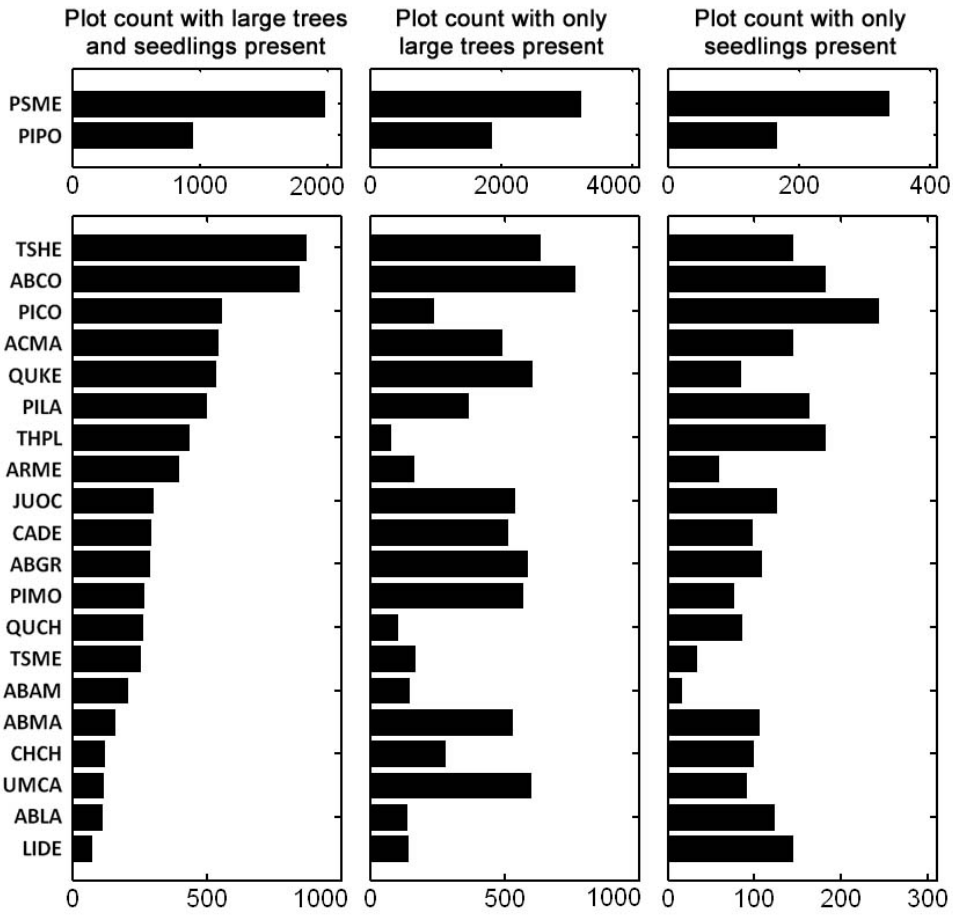


Figure 6. Box-plots of model fits for single-predictor models by climate variables and life stage. Only models with a model fit or AUC greater than or equal to 0.75 are included. AUC ranges from 0 to 1 where 0.5 and below represents a model no better achieved from random chance alone. Pie charts below each box-plot represent the proportion of models with the best predictor as either a mean (grey) or an extreme (white). The number of models corresponding to each box-plot are shown across the top. Seedling models are denoted with 'SD' and adult models are denoted with 'LG.'

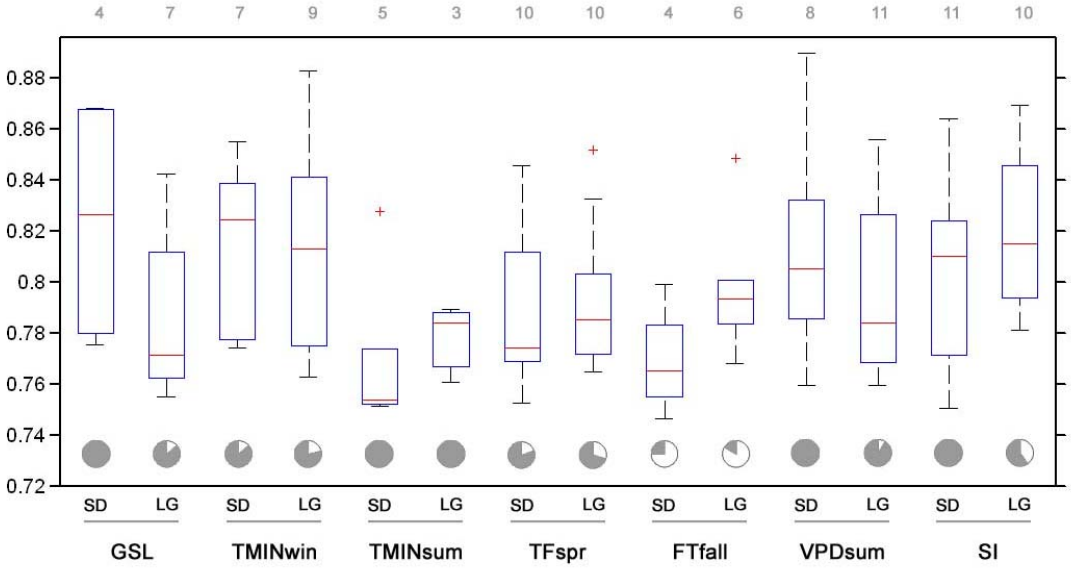


Figure 7. Predicted values for adult (blue solid line) and seedling (red dashed line) probability of occurrence relative to the four climate variables with the highest proportional representation of extremes: TMINwin, TFspr, FTfall, and SI (see Fig. 6). Presence/absence data were force-fitted to long-term means, and the species where seedling and adults models had AUCs greater than or equal to 0.75 were plotted here. One species and climate variable is an exception to this, LIDE for GSL (top left). This was included to complement data presented in Fig. 8. Probability of occurrence data were first standardized by their maximum values to compensate for different probabilities of detection (or FIA plot sizes) for seedling and adults.

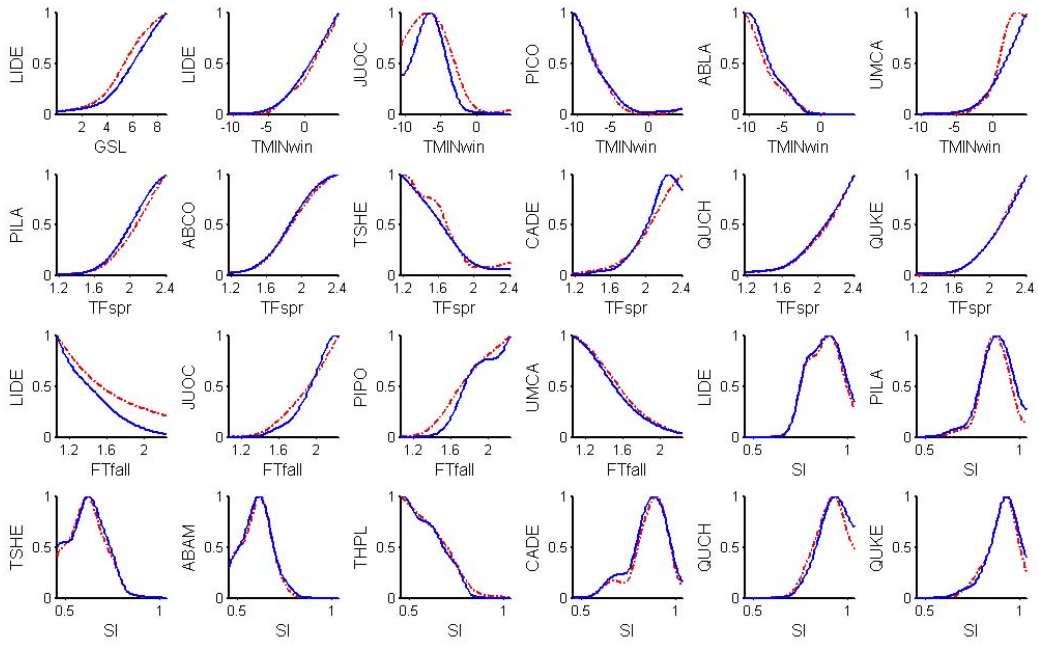
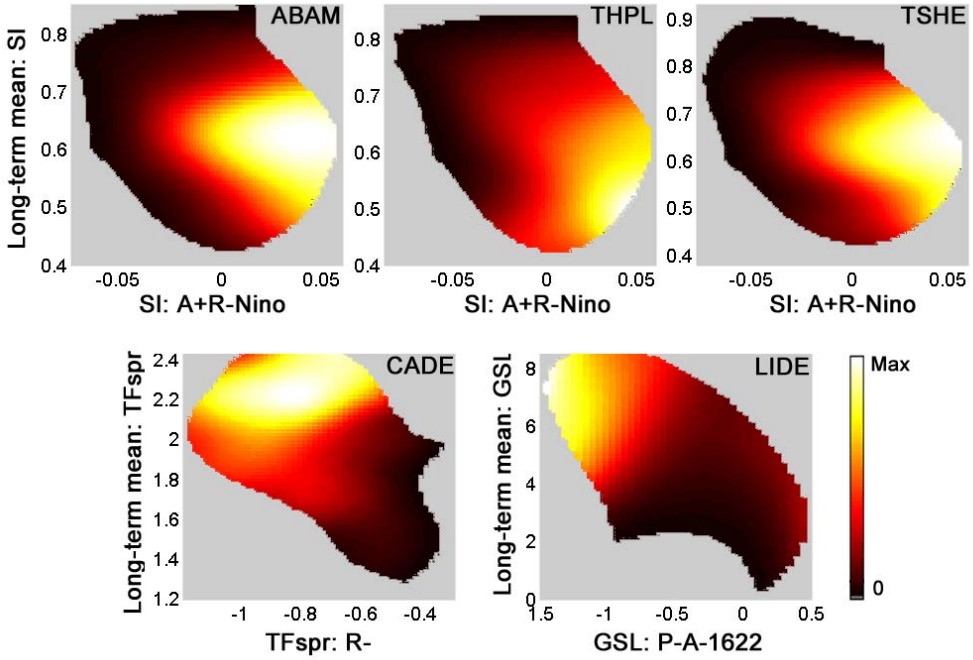


Figure 8. Multi-factor response surfaces are shown as bird's eye views of species' probability of occurrence versus a long-term mean and an extreme. Predicted values are shown for five species based on the criteria that a long-term term mean for a seedling niche was selected while an extreme was selected for the adult niche. Also, model fits were greater than or equal to an AUC of 0.75. All surfaces are adult responses. In each case except for ABAM, the best two-predictor model for the adults was the extreme and the long-term mean. For ABAM, the best three-predictor models contained the mean. Below is the two-predictor model for ABAM that includes the mean. The top row consists of response surfaces based on the same *x* and *y* as labelled, and default settings were used in NPMR to avoid extrapolating in areas of little data. Grey areas denote where no predictions were made. To view adult and seedling responses for these five species when forced to the long-term mean alone see Figure 7. The model fits (as AUCs) for species from left-to-right and top-to-bottom are equal to 0.88, 0.84, 0.88, 0.87, and 0.88.



TABLES

Table 1. Twenty-two species were studied. Species codes, and Latin and common names are shown.

Code	Latin name	Common name
ABAM	*Abies amabilis*	Pacific silver fir
ABCO	*Abies concolor*	white fir
ABGR	*Abies grandis*	grand fir
ABLA	*Abies lasiocarpa*	subalpine fir
ABMA	*Abies magnifica*	California red fir
ACMA	*Acer macrophyllum*	bigleaf maple
ARME	*Arbutus menziesii*	Pacific madrone
CADE	*Calocedrus decurrens*	incense cedar
CHCH	*Castanopsis chrysophylla*	golden chinkapin
JUOC	*Juniperus occidentalis*	western juniper
LIDE	*Lithocarpus densiflorus*	tanoak
PICO	*Pinus contorta*	lodgepole pine
PILA	*Pinus lambertiana*	sugar pine
PIMO	*Pinus monticola*	western white pine
PIPO	*Pinus ponderosa*	ponderosa pine
PSME	*Pseudotsuga menziesii*	Douglas fir
QUCH	*Quercus chrysolepis*	canyon live oak
QUKE	*Quercus kelloggii*	California black oak
THPL	*Thuja plicata*	western red cedar
TSHE	*Tsuga heterophylla*	western hemlock
TSME	*Tsuga mertensiana*	mountain hemlock
UMCA	*Umbellularia californica*	California laurel

Table 2. Seven climate variables were derived. Derivations and justifications for variables are shown.

	Derivation	**Justification**
Growing Season Length (GSL)	Number of months in a year where the mean monthly minimum temperature exceeds 5^{o}C.	Growing season length affects growth, reproduction, dormancy, and the timing of budburst, budset, flowering, and seedset (Kozlowski et al. 1991). Growing season length affects the supply of energy for metabolism and photosynthesis (Larcher 2003). Longer growing seasons may lead to more growth or competitive ability without drought (Myneni et al. 1997, Zhao and Runing 2010, Loehle 1998). Rapid change in length of growing season can foster range expansion for species with suited phenologies (Holmgren et al. 2006).
Winter Minimum Temperature (TMINwin)	Mean minimum monthly temperature averaged across November, December, and January for a year.	Cold temperatures can cause tissue damage and mortality from ice nucleation, cell membrane rupture, and cellular dehydration (Woodward 1987, Raison et al. 1979, Becwar and Burke 1982, Levitt 1980).
Summer Minimum Temperature (TMINsum)	Mean minimum monthly temperature averaged across June, July, and August for a year.	Plants are more sensitive to detrimental effects of cold temperatures during the the growing season (Lyons et al. 1979). Photo-inhibition and oxidation caused by low summer temperatures combined with high light intensity can kill seedlings (Germino and Smith 1999, Ball et al. 1991).

Table 2 Continued.

Name	Derivation	Justification
Spring Thaw-Freeze (TFspr)	The rate of change was calculated among monthly maximum temperature anomalies for two monthly transitions: March to April, and April to May. Anomalies were calculated as deviations from the long-term mean for that month (reference period of 108 years). The April anomaly was subtracted from the March anomaly, and the May anomaly was subtracted from the April anomaly. The minimum rate of change was chosen among the two transitions to get the maximum drop from warm anomaly to cold anomaly for the season. This was then multiplied by negative one so that the degree of spring thaw-freeze would increase with the index value.	Late spring frost preceded by early warming sensitizes plant tissues to the negative effects of cold temperatures (see above) (Gu et al. 2008, Augspurger 2009, Neilson and Wullstein 1983, Zon 1904). Plant species differentially utilize nitrogen pulses associated with freeze-thaw cycles (Bilbrough and Caldwell 1997). Freeze-thaw events cause xylem embolism, which can impair physiological processes and lead to hydraulic failure (Sperry and Sullivan 1992).
Fall Freeze-Thaw (FTfall)	Same as above except the monthly transitions were: September to October, and October to November. The most positive rate of change was chosen among the two transitions for fall to get the maximum jump from cold anomaly to warm anomaly for the season.	Cold acclimation occurs in the fall and is induced by cool non-freezing temperatures (Levitt 1980). Late warming events can interfere with cold acclimation and sensitize plants to the detrimental effects of cold temperatures occurring in winter. Freeze-thaw events also cause xylem embolism (Sperry and Sullivan 1992).

Table 2 Continued.

Name	Derivation	Justification
Summer Vapor Pressure Deficit (VPDsum)	The mean monthly VPD was averaged for June, July, and August. Mean monthly VPD was derived as $VPD = e_s - e$, where e_s is the mean monthly saturation vapor pressure and e is the mean monthly vapor pressure. Both e_s and e were based on mean monthly temperatures (dew point and air) (Murray 1967). We used the sine form weighted mean daily air temperature to derive monthly mean air temperature. This allows for accurate estimation of mean monthly temperature from mean monthly minimums and maximums (Goeckede, personal communication, Thornton et al. 1997). Grids of dew point temperature (T_s) supplied from PRISM are based on hourly measurements over a 24 hour period and did not require a transformation.	Vapor pressure deficit largely drives plant transpiration. The control of transpiration in response to vapor pressure deficit differs by species (Oren et al 2002). Excessive transpiration during drought can lead to dehydration, hydraulic failure, and mortality depending on stomatal response (McDowell et al. 2008). However, efficiency of transpiration can endow competitive advantage (Bunce et al. 1977, Sperry et al. 2008). Low vapor pressure deficit and seasonal fog occur in our study region and can decouple tree hydration from soil water deficit by direct water uptake from leaf surfaces (Burgess and Dawson 2004).
Seasonality of Precipitation (SI)	The Seasonality Index measures the departure from uniform precipitation (Walsh and Lawler 1981). SI equals zero when there is uniform precipitation across 12 months.	Changes in the timing of precipitation alter germination, phenology (or timing of events like flowering), recruitment, and survival (Weltzin et al. 2003, Penuelas et al. 2004, Kimball et al. 2010, Franks et al. 2007, Neilson and Wullstein 1986).

Table 3. Five indices of climatic oscillations were used to classify years into different groups. Groups contributed to 'climate schemes.' The indices and their classification rules are described. The time period from 1976 until 2007 was also included to represent give account to anthropogenic climate change.

Name	Definition	Classification Rules	Data Source
Pacific Decadal Oscillation (PDO)	The leading principal component of monthly anomalies in sea surface temperatures above 20N in North Pacific Ocean (Zhang et al. 1997, Mantua et al. 1997).	PDO: Years are considered positive PDO if the annual mean for the index is positive. The converse is true for negative. We defined intervals across years through dominant behavior of annual means. High frequency variation was minor and ignored.	Joint Institute for the Study of the Atmosphere and the Ocean, Climate Impacts Group, University of Washington
Atlantic Multi-decadal Oscillation (AMO)	Sea surface temperature anomalies in the Atlantic north of the equator. The index is normalized to unit standard deviation with monthly temporal resolution. The index was smoothed using a ten-year running mean following Enfield et al. (2001).	AMO: Years are considered positive AMO if the annual mean for the index is positive. The converse is true for negative.	National Oceanic and Atmospheric Association
Trans-Niño Index (TNI)	The difference between normalized sea surface temperature anomalies among equatorial eastern and equatorial west-central Pacific. The index discerns the type of El Niño or La Niña by measuring the gradient preceding or following an event. Prior to 1976, the TNI of "opposite sign" occurs 3 to 12 months after an event (defined by Niño 3.4 Index) and TNI of the same sign occurs before. This relationship is reversed after 1976 (Trenberth and Stepaniak 2001). The index is normalized to unit standard deviation with monthly temporal resolution.	TNI75: Extreme events had annual mean index values above 0.75^{o}C or below -0.75^{o}C. TNI4: Events had annual mean index values above 0.4^{o}C or below -0.4^{o}C.	National Center for Atmospheric Research

Table 3 Continued.

Name	Definition	Classification Rules	Data Source
Niño 3.4 Index (N3.4)	Sea surface temperature anomalies in the Niño 3.4 region (central equatorial Pacific). The index is normalized to unit standard deviation with monthly temporal resolution.	N3475: Extreme events had annual mean index values above 0.75^{o}C or below -0.75^{o}C. N344: Events had annual mean index values above 0.4^{o}C or below -0.4^{o}C.	National Center for Atmospheric Research
Southern Oscillation Index (SOI)	The difference of monthly mean sea level *pressure* anomalies among Tahiti (in the central pacific) and Darwin (in Australia or eastern Pacific).The anomalies are normalized to their standard deviate with monthly temporal resolution. A low-pass filter is then used to calculate the index to increase the signal to noise ratio (Trenberth 1984).	SOI75: Extreme events had annual mean index values above 0.75^{o}C or below -0.75^{o}C. SOI4: Events had annual mean index values above 0.4^{o}C or below -0.4^{o}C.	National Center for Atmospheric Research
R1	The time period from 1976 to 2007 representing occurrence of anthropogenic climate change. This is not a recognized climate oscillation but rather a unique temporal period to be given account.	Not applicable as no index was used	Not applicable

Table 4. Different oscillation schemes were tested and compared as temporal grouping variables. The schemes contained all combinations of different oscillations (see Table 3 for code definitions). All combinations containing ENSO indices shown were repeated for ENSO indices with a lower cut-off (plus or minus 0.4^{o}C, Table 3). The codes in this table all represent ENSO indices with a cut-off of plus or minus 0.75^{o}C. Hence, codes containing ENSO end in '75.'

Code for oscillation scheme	Oscillations involved (explicit definitions in Table 3)
Decadal+ scale	
PDO	Pacific Decadal Oscillation
AMO	Atlantic Multi-decadal Oscillation
R1	Anthropogenic climate change
PDOgr	PDO as unique groups through time
AMOgr	AMO as unique groups through time
PDOAMO	PDO, AMO
PDOAMOgr	Intersection of PDO and AMO as unique groups through time
AMOR	AMO, R1
PDOR	PDO, R1
PDOAMOR	PDO, AMO,R1
AMOgrR	AMO as unique groups through time intersected with R1
ENSO alone, strong events	
TNI75	Trans-Niño
SOI75	Southern Oscillation Index
N3475	Niño 3.4
ENSO, strong events, together with decadal+ scale oscillations	
PTNI75	PDO, TNI75
PSOI75	PDO, SOI75
PN3475	PDO, N34
PRTNI75	PDO, R1, TNI75
PRSOI75	PDO, R1, SOI75
PRN3475	PDO, R1, N3475
ATNI75	AMO, TNI75
ASOI75	AMO, SOI75
AN3475	AMO, N3475
ARTNI75	AMO, R1, TNI75
ARSOI75	AMO, R1, SOI75
ARN3475	AMO, R1, N3475
RTNI75	R1, TNI75
RSOI75	R1, SOI75
RN3475	R1, N3475
PATNI75	PDO, AMO, TNI75
PASOI75	PDO, AMO, SOI75
PAN3475	PDO, AMO, N3475
PARTNI75	PDO, AMO, R1, TNI75
PARSOI75	PDO, AMO, R1, SOI75
PARN3475	PDO, AMO, R1, N3475
All schemes above were repeated for moderate ENSO events	

Table 5. MRPP results for winning oscillation schemes by climate variable. *A*-values and *P*-values are reported where *A*-values are the effect size or the degree of separation among groups. *A*=0 for random expectation and has an upper bound at 1, meaning that all values are identical within groups and different among groups. An *A*-value of 0.1 represents a strong degree of separation. Oscillation scheme codes are explained in Table 4.

Climate variable	Best oscillation scheme	Oscillations involved	*A*-value	*P*-value
GSL	PAgr	PDO and AMO phases as unique groups through time	0.12	<0.001
TMINwin	PARSOI4	PDO, AMO, R1, SOI4	0.21	<0.001
TMINsum	PAgr	PDO and AMO phases as unique groups through time	0.24	<0.001
TFspr	RSOI4	R1, SOI4	0.02	0.021
FTfall	ARN344	AMO, R1, N344	0.05	0.001
VPDsum	PARTNI75	PDO, AMO, R1, TNI75	0.15	<0.001
SI	ARSOI75	PDO, AMO, SOI75	0.02	0.025

Table 6. Eigenvalues and variance explained by Principal Components Analysis (PCA) performed separately for three groups of variables: all climate extremes, all means combined with extremes, and all means. Values correspond to the first three principal components. The sample for each analysis represented FIA plots used in model building or forested land in Oregon, California, and Washington (N=2500). The PCA was based on a matrix of correlation coefficients among the data.

Axis	Eigenvalue	Variance explained, %
Just extremes		
Axis 1	12.635	15.409
Axis 2	9.285	11.323
Axis 3	7.223	8.808
Means and extremes		
Axis 1	13.835	15.544
Axis 2	10.716	12.040
Axis 3	8.046	9.040
Just means		
Axis 1	3.523	50.325
Axis 2	2.316	33.089
Axis 3	0.645	9.221

Table 7. Two by two tables of count data for Fischer Exact Tests and their respective P-values. Numbers of species are counted with the best single predictor models choosing the long-term mean (mean+) versus an extreme (extreme+). Counts were made for all models (above) and again for a subset of models just with good fits (or AUCs greater than 0.75)(below). The starting pool for counting was 22 species' models multiplied by seven climate variables and two life stages (bottom table). Fischer's Exact Tests were performed separately on each two by two count table to yield individual *P*-values.

	Mean+	Extreme+	P-value
All best single predictor models			
Large trees	94	60	
Seedlings	118	36	0.004
All best single predictor models with good fits			
Large trees	40	16	
Seedlings	43	6	0.054

Table 8. Input and results test the probability of a result as extreme as or more extreme than the number of long-term means chosen per climate variable and by life stage. 'Mean+' represents number of species where the long-term mean was selected over an extreme. Left column shows 'Mean+' for all models where the total per climate variable is 22. *P*-values are the next column right. Numbers of predictors per climate variable are found in the right-most column.

	Num species with Mean+ (from all species)	P-value	# of preds
Seedlings			
GSL	22	<0.0001	3
TMINwin	17	<0.0001	7
TMINsum	21	<0.0001	3
TFspr	19	<0.0001	4
FTfall	7	0.0234	7
VPDsum	16	<0.0001	7
SI	16	<0.0001	7
Large trees			
GSL	19	<0.0001	9
TMINwin	15	<0.0001	19
TMINsum	17	<0.0001	9
TFspr	14	<0.0001	7
FTfall	5	0.0129	13
VPDsum	15	<0.0001	19
SI	9	<0.0001	13

Chapter 5. Conclusions

This dissertation contributes to the study of ecological complexity by making methodological strides in analysis of patterns of species' occurrence. Philosophers of science posit that understanding of complex phenomena comes through one of two ways. It can come through pattern analysis of emergent phenomena (von Hayek 1964). It may also come through detailed data acquisition and multi-scalar modeling of attributes of a system (which is unbounded) and its interactions, feedbacks, and non-linear restructuring of relationships through time (von Hayek 1964, Taylor 2005). However, the latter approach may be impractical and exceedingly difficult if not impossible to achieve for some problems (von Hayek 1964, Taylor 2005). Species' occurrence data are composed of individuals and populations that vary genetically, occur at different points in their life cycle, and are embedded within varied geographic, environmental, and biological contexts; in turn, a species and its constituent parts also contribute to those contexts. Tool development that can address this type of complexity is warranted in ecology. The work presented here helps to move us toward better recognition and characterization of complexity, specifically with respect to the species' niche.

I discuss the respective contribution of each chapter below. Future research directions that emerge from each chapter follow the conclusive summary for that chapter. Overall future research directions offered by the entire body of work are found at the end.

Chapter 2 developed the first tools for multi-dimensional threshold analysis in state space. I derived threshold strength and diagonality, two indices of shape attributes, and showed they can be quantified from ecological response surfaces. I also demonstrated that these tools can allow us to ask questions about what drives thresholds including for species' niche models. Also, for the purpose of threshold method development, different data mining methods were tested for their ability to retrieve the shape of a response surface. Non-parametric Multiplicative Regression (NPMR), a kernel regression technique designed to automatically accommodate interactions occurring at any scale, outperformed other methods. Results from Chapter 2 improved our ability to

retrieve the shape of ecological response surfaces and accept response surfaces as emergent snapshots of complex systems. Chapter 2 helped to answer the call for improvement in methods of shape detection in ecological responses (Oksanen and Minchin 2002, Soberón and Nakamura 2009). The need to understand model bias for data mining methods has also been identified but poorly explored (Elith and Graham 2009). Chapter 2 examined the cause and differential expression of model bias for common data mining methods. The bias manifested as geometric constraints imposed by the algorithms.

Chapter 2 sets the stage for future tests of the performance of other data mining approaches. Techniques of particular interest include Multivariate Adaptive Regression Splines (MARS) and Boosted Regression Trees (BRT) as they perform well in tests of niche modeling using real ecological data (Elith et al. 2008, Elith et al 2006). Another enticing future topic would be the development of a roving window method that can measure threshold strength and diagonality at various scales within a response surface. Along this line, a title for a future publication could be, “The development of a threshold-seeking automaton in multidimensional state space.” Another publication could also explore extension of the threshold strength and diagonality indices to more than three dimensions of state space (or *n*-dimensional state space). Given the inherent complexity of ecosystems, empirical characterization of thresholds in more than three dimensions is warranted (Limburg et al. 2002). Additionally, results not reported (but associated with analyses from Chapter 4) reveal that shapes of species’ occurrence relative to climate extremes exhibit what appear to be thresholds. Their measurement and study is warranted in a future paper. Finally, upon their application, tools from Chapter 2 can spur many different hypotheses for further testing depending on the system and scale of study.

Chapter 3 addressed an important methodological challenge common in niche modeling: the effect of combining data collected by different methods (Barry and Elith 2006, Elith et al. 2002). The following question was asked, do data from two different collection methods or forest inventories yield the same niche models? Our conclusions revealed that there is a small effect of combining data from different forest inventories.

Further, the fit of all the models was notably high across data sets suggesting that tree species' occurrence is strongly linked to climate in the Pacific coastal United States. Also, results showed little evidence for changes in niche structure of species' occurrence in relation to climate across two decades. Finally, results prompted further interpretive investigation associated with model selection in NPMR. It seemed that cross-validation was a more robust method of model selection compared to external validation for the large FIA data sets.

The most immediately and intriguing future question emerging from Chapter 3 is how to measure temporal change over the past two decades given what we know about the effect of inventory type. One possible way would be to examine differences among geographic projections made from niche models (e.g. Chapter 3 but with more difference maps). Niche models are convenient in this context because climatic redundancy can dampen sampling differences. This assumes that tree species' niches are not changing across two decades, which is a research question in its own right. Another possibility for trend monitoring in the future would be to use NPMR instead for direct spatial interpolation of species' occurrence data, without reference to climate, then examine differences through time by subtracting spatially-interpolated maps. A comparison of niche modeling versus spatial interpolation in trend monitoring is a question worthy of pursuit. Tracking of species' occurrence across the landscape through time is an important future step given that we are in the midst of a global, unplanned experiment of anthropogenic climate change. Forecasts reveal that biomes and plant species' ranges are changing and will continue to change (e.g. Walther et al. 2002, Bachelet et al 2001). Improved methods for monitoring will establish a strong baseline for better understanding species' responses through time. If we can harness the information in the FIA database despite political and financial setbacks that alter funds from year to year to affect sampling, we can explore questions associated with niche dynamics in the face of environmental change.

Chapter 4 tested for empirical evidence to support a proposed hypothesis behind tree species' distributions: tree species' niches can respond more strongly to climate

extremes than to means, and this differs by life stage. An unconventional way of characterizing climate interactions to quantify climate extremes was developed to test this hypothesis. The measurement and separation of the climate change signal among a backdrop of natural variability is an area of active research (Easterling et al. 2000, Wikle 2003), and here, complex climate interactions among oscillations were characterized while keeping track of climate change effects. The approach accommodated temporal interactions at different scales. Additionally, two new indices of freeze-thaw events occurring in the spring and fall were developed for climate time series to quantify such events for the purpose of this study. A subset of species' niche models was found to have more variability explained by extremes than means. However, seedling niches were not affected as often by climate extremes compared to adults. The reasons for juveniles being "less sensitive" are unclear. The most compelling explanation was that adult niches apparently had a longer "memory" to climatic events and thus a higher likelihood of being shaped by climate extremes. Finally, climate means still won over extremes for the majority of species, supporting the logic that has sustained niche modeling for decades. Overall, the results highlight a grand challenge in forecasting tree species' occurrence. A subset of tree species are affected by extremes modulated by decadal-scale climate variation, yet long-term climate forecasts are made from models that cannot capture decadal-scale climate variation. Thus, the climate forecasts that many forest scientists and managers rely on to make ecological projections are missing an important piece of the puzzle.

Chapter 4 inspires many future research questions. First, why are seedling niches on a whole less sensitive to climate extremes defined by climate oscillations? Second, do the patterns of winning oscillation schemes for the different climate variables hold for weather station data in addition to PRISM data? Third, what do multi-factor response surfaces look like for tree species' occurrence as a function of climate extremes? Fourth, how can we better conceptualize, accommodate, quantify, and understand influential niche factors fluctuating in space and time?

In grand summary, this dissertation developed tools to study methodological and ecological topics associated with species' distribution patterns as complex systems in ecology. Particular emphasis was devoted to unconventional measurement of thresholds, climate interactions, climate extremes, data mining methods, species' niches, and issues associated with large-scale biological data integration. The themes established from these works can be developed in many ways. For example, are ecological thresholds exhibited by species' niches with respect to climate extremes? Is there commonality in processes that contribute to those thresholds? If thresholds occur for species' niches, does mortality peak in correspondence to climate domains where thresholds occur? Also, do shape features of response surfaces in niche space change for different snapshots in time? These questions exemplify how the shapes of response surfaces can be better exploited to probe processes behind species' niches while at the same time integrating the topics of thresholds and climate extremes.

Understanding and describing the shape of multidimensional response surfaces has applications well beyond modeling the niche of a species. For example, response surfaces measured for any real data set can be compared to response surfaces produced by dynamical computer models. The comparison of shape can help verify that the emergent behavior of a computer model matches that of the natural system. Understanding the behavior of complex computer models is a topic of active research in the study and prediction of complex systems. This would be an intriguing research topic to bring to ecology and global change science.

Bibliography

Allen, C.D., Breshears, D.D. 1998. Drought-induced shift of a forest-woodland ecotone: rapid landscape response to climate variation. Proceedings of the National Academy of Sciences, U.S.A. 95: 14839-14842.

Allen, T.F.H., Starr, T.B. 1982. Hierarchy: perspectives for ecological complexity. University of Chicago, Chicago.

Andersen T., Cartensen, J., Hernández-Garcia, E., Duarte, C.M. 2009. Ecological thresholds and regime shifts: approaches to identification. Trends in Ecology and Evolution 24: 49-57.

Araújo, M.B., Pearson, R.G., Thuiller, W., Erhard, M. 2005. Validation of species-climate impact models under climate change. Global Change Biology 11: 1504-1513.

Araújo, M.B., Guisan, A. 2006. Five (or so) challenges for species' distribution modeling. Journal of Biogeography 33: 1677-1688.

Austin, M.P., Nicholls, A.O., Margules, C.R. 1990. Measurement of the realized qualitative niche: environmental niches of five Eucalyptus species. Ecological Monographs 60: 161-177.

Austin, M.P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. Ecological Modelling 157: 101-118.

Austin, M. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecological Modelling 200: 1-19.

Augspurger, C.K. 2009. Spring 2007 warmth and frost: phenology, damage and refoliation in a temperate deciduous forest. Functional Ecology 23: 1031-1039.

Azuma, D.L., Dunham, P.A., Hiserote, B.A., Veneklase, C.A. 2002. Timber resource statistics for eastern Oregon, 1999. USDA Forest Service Pacific Northwest Research Station Resource Bulletin PNW-RB-238.

Azuma, D.L., Bednar, L.F., Hiserote, B.A., Veneklase, C.A. 2004. Timber resource statistics for western Oregon, 1997. USDA Forest Service Pacific Northwest Research Station Resource Bulletin PNW-RB-237: revised.

Bachelet, D., Neilson, R.P., Lenihan, J.M., Drapek, R.J. 2001. Climate change effects on vegetation distribution and carbon budget in the United States. Ecosystems 4: 164-185.

Baker, M.E., King, R.S. 2010. A new method for detecting and interpreting biodiversity and ecological community thresholds. Methods in Ecology and Evolution 1: 25-37.

Ball, M.C., Hodges, V.S., Laughlin, G.P. 1991. Cold-induced photoinhibition limits regeneration of snow gum at tree line. Functional Ecology 5: 663-668.

Barrett, T.M. 2004. Estimation Procedures for the Combined 1990s Periodic Forest Inventories of California, Oregon, and Washington. USDA Forest Service, Pacific Northwest Research Station General Technical Report PNW-GTR-597. Portland, Oregon.

Barry, S., Elith, J. 2006. Error and uncertainty in habitat models. Journal of Applied Ecology 43: 413-423.

Bartlein, P.J., Prentice, I.C., Webb, T. 1986. Climatic response surfaces from pollen data for some eastern North American taxa. Journal of Biogeography 13: 35-57.

Bechtold, W.A., Patterson, P.L. 2005. The enhanced Forest Inventory and Analysis Program--national sampling design and estimation procedures. USDA Forest Service, Southern Research Station General Technical Report SRS-80. Asheville, North Carolina.

Becwar, M.R., Burke, M.J. 1982.Winter hardiness associations and physiography of woody timberline flora. In: Li, P.H., Sakai, A., eds. Plant Cold Hardiness and freezing stress: mechanisms and crop implications, Vol. 2. Pg. 307-323. Academic press, New York.

Benson, L., Braddock, L., Smoot, J., Mensing, S., Lund, S., Stine, S., Sarna-Wojcicki, A. 2003. Influence of the Pacific Decadal Oscillation on the climate of the Sierra Nevada, California and Nevada. Quaternary Research 59: 151-159.

Berk, R.A. 2006. An introduction to ensemble methods for data analysis. Sociological Methods Research 34: 263-295.

Berryman, A.A. 1982. Biological control, thresholds and pest outbreaks. Environmental Entomology 11: 544-549.

Bilbrough, C.J., Caldwell, M.M. 1997. Exploitation of springtime ephemeral N pulses by six great basin plant species. Ecology 78: 231-243.

Bitterlich, W. 1948. Die Winkelzählprobe. Allgemeine Forst-und Holzwirtschaftliche Zeitung 59: 4-5.

Boyce, M.S., Haridas, C.V., Lee, C.T., National Center for Ecological Analysis and Synthesis (NCEAS) Stochastic Demography Working Group. 2006. Demography in an increasingly variable world. Trends in Ecology and Evolution 21: 141-148.

Bowersox, M.A., Brown, D.G. 2001. Measuring the abruptness of patchy ecotones -- a simulation-based comparison of landscape pattern statistics. Plant Ecology 156: 89-103.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. 1984. Classification and Regression Trees. Chapman and Hall, New York.

Breiman, L. 2001. Random forests. Machine Learning 45: 5-32.

Brenden, T.O., Wang, L., Su, Z. 2008. Quantitative identification of disturbance thresholds in support of aquatic resource management. Environmental Management 42: 821-832.

Bunce, J.A., Miller, L.N., Chabot, B.F. 1977. Competitive exploitation of soil water by five eastern North American tree species. Botanical Gazette 138: 168-173.

Burgess, S.S.O., Dawson, T.E. 2004. The contribution of fog to the water relations of *Sequoia sempervirens* (D. Don): foliar uptake and prevention of dehydration. Plant, Cell and the Environment 27: 1023-1034.

Cadenasso, M. L., Pickett, S.T.A., Weathers, K.C., Jones, C.G. 2003. A Framework for a theory of ecological boundaries. Bioscience 53: 750-758.

Cayan, D., Dettinger, M., Diaz, H., Graham, N. 1998. Decadal variability of precipitation over western North America. Journal of Climate 11: 3148-3166.

Chambers, J.M., Hastie, T.J. 1992. Statistical Models in S. Wadsworth and Brooks/Cole, California.

Chambers, J.M., Cleveland, W.S., Kleiner, B., Tukey, P.A. 1983. Graphical Methods for Data Analysis. Wadsworth Publishing Company, California.

Chase, J.M., Leibold, M.A. 2001. Ecological Niches: linking classical and contemporary approaches. University of Chicago Press, Chicago.

Chaubert-Pereira, F., Caraglio, Y., Lavergne, C., Guédon, Y. 2009. Identifying ontogenetic, enironmental, and individual components of forest tree growth. Annals of Botany 104: 883-896.

Christiansen E., Waring, R.H., Berryman, A.A. 1987. Resistance of conifers to bark beetle attack: searching for general relationships. Forest Ecology and Management 22: 89-106.

Chuine, I., Beaubien, E. 2001. Phenology is a major determinant of temperate tree range. Ecology Letters 4: 500-510.

Colwell, R.W., Rangel, T.F. 2009. Hutchinson's duality the once and future niche. Proceedings of the National Academy of Sciences, U.S.A. 106: 19651-19658.

Critchfield, W.B. 1988. Hybridization of the California firs. Forest Science 34: 139-15.

Crossman, N.D., Bryan, B.A. , Cooke, D.A. 2011. An invasive plant and climate change threat index for weed risk management: Integrating habitat distribution pattern and dispersal process. Ecological Indicators 11: 183-198.

Daly, C., Gibson, W.P., Taylor, G.H., Johnson, G.L., Pasteris, P. 2002. A knowledge-based approach to the statistical mapping of climate. Climate Research 22: 99-113.

Damgaard, C. 2006. Modelling ecological presence-absence data along an environmental gradient: threshold levels of the environment. Environmental and Ecological Statististics 13: 229-236.

Davis, M.B., Shaw, R.G. 2001. Range shifts and adaptive responses to Quaternary climate change. Science 292: 673-679.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kuhn, I., Ohlemuller, R., Peres-Neto, P., Reineking, B., Schroder, B., Schurr, F.M., Wilson, R. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30: 609-628.

Easterling, D.R., Meehl, G.A., Parmesan, C., Changon S.A., Karl, T.R., Mearns, L.O. 2000. Climate extremes: observations, modeling, and impacts. Science 289: 2068-2074.

Efron, B., Tibshirani, R. 1991. Statistical data analysis in the computer age. Science 253: 390-395.

Elith, J., Burgman, M.A., Regan, H.M., 2002. Mapping epistemic uncertainties and vague concepts in predictions of species distribution. Ecological Modelling 157: 313-329.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettman, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T, Phillips, S., Richardson, K., Pereira, R.S., Schapire, R.E., Soberon, J., Williams, S.E., Wisz, M., Zimmermann, N.E. 2006. Novel methods improve predictions of species' distributions from occurrence data. Ecography 29: 129-151.

Elith, J., Leathwick, J.R., Hastie, T. 2008. A working guide to boosted regression trees. Journal of Animal Ecology 77: 802-813.

Elith, J., Graham, C.H. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography 32: 66-77.

Enfield, D.B., Mestaz-Nunez, A.M., Trimble, P.J. 2001. The Atlantic Multi-decadal Oscillation and its relation to rainfall and river flows in the continental U.S. Geophysical Research Letters 28: 2077-2080.

Engelbrecht, B.M.J., Comita, L.S., Condit, R., Kursar, T.A., Tyree, M.T., Turner, B.L., Hubbell, S.P. 2007. Drought sensitivity shapes species distribution patterns in tropical forests. Nature 447: 80-82.

Érdi, P. 2008. Complexity Explained. Springer, Germany.

European Commission. 1997. Study on European forestry information and communication system: Report on forest inventory and survey systems. European Commission, Luxembourg.

Evans, J.S., Cushman, S.A. 2009. Gradient modeling of conifer species using random forests. Landscape Ecology 24: 673-683.

Fawcett, T. 2006. An introduction to ROC analysis. Pattern Recognition Letters 27: 861-874.

Ferrell, G.T., Smith, R.S. 1976. Indicators of *Fomes annosus* root decay and bark beetle susceptibility in sapling white fir. Forest Science 22: 365-369.

Franks, S.J., Sim, S., Weis, A.E. 2007. Rapid evolution of flowering time by an annual plant in response to climate fluctuation. Proceedings of the National Academy of Sciences, U.S.A. 104: 1272-1282.

Fortin, M. 1994. Edge detection algorithms for two-dimensional ecological data. Ecology 75: 956-965.

Fortin, M., Dale, M.R.T. 2009. Spatial autocorrelation in ecological studies: A legacy of solutions and myths. Geographical Analysis 41: 392-397.

Gause, G.F. 1936. The struggle for existence. Williams and Wilkins, Baltimore.

Germino, M.J., Smith, W.K. 1999. Sky exposure, crown architecture, and low temperature photoinhibition in conifer seedlings at alpine treeline. Plant, Cell and the Environment 22: 407-415.

Goeckede, Mathias. Research Associate, TERRA-PNW. College of Forestry, Oregon State University. Corvallis, Oregon.

Goldenfield, N., Kadanoff, L.P. 1999. Simple lessons from complexity. Science 284: 87-89.

Gomulkiewicz, R., Holt, R.D. 1995. When does evolution by natural selection prevent extinction? Evolution 49: 201-207.

Graham, C., Ron, S.R., Santos, J.C., Schneider, C.J., Mortiz, C. 2004. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. Evolution 58: 1781-1793.

Gray, A.N. 2003. Monitoring stand structure in mature coastal Douglas-fir forests: effect of plot size. Forest Ecology and Management 175: 1-16.

Gray, S.T., Betancourt, J.L., Fastie, C.L., Jackson, S.T. 2003. Patterns and sources of multi-decadal oscillations in drought sensitive tree-ring records from the central and southern Rocky Mountains. Geophysical Research Letters 30: 2002GL01654.

Grinnell, J. 1917. The niche relationships of the California thrasher. Auk. 34: 427-433.

Groffman, P.M., Baron, J.S., Blett, T., Gold, A.J., Goodman, I., Gunderson, L.H., Levinson, B.M., Palmer, M.A., Paerl, H.W., Peterson, G.D., Poff, N.L., Rejeski, D.W., Reynolds, J.F., Turner, M.G., Weathers, K.C., Wiens, J. 2006. Ecological thresholds: the key to successful environmental management or an important concept with no practical application? Ecosystems 9: 1-13.

Grosenbaugh, L.R. 1952. Plotless timber estimates—new, fast, easy. Journal of Forestry 50: 32-37.

Grosenbaugh, L.R., Stover, W.F. 1957. Point sampling compared with plot sampling in southeast Texas. Forest Science 3: 2-14.

Grubb, P.J. 1977. The maintenance of species richness in plant communities: the importance of the regeneration niche. Biological Reviews 52: 107-145.

Gu, L., Hanson, P.J., Post, W.M., Kaiser, D.P., Yang, B., Nemani, R., Pallardy, S.G., Meyers, T. 2008. The 2007 Eastern US spring freeze: increased cold damage in a warming world? Bioscience 58: 253-262.

Guisan, A., Zimmerman, M.E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135: 147-186.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overont, J., Aspinall, R., and Hastie, T. 2006. Making better biogeographical predictions of species' distributions. Journal of Applied Ecology 43: 386-392.

Guisan, A., Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S.J., Townsend Peterson, A. 2007. What matters for predicting the occurences of trees: Techniques, data, or species'characteristics? Ecological Monographs 77: 615-630.

Gutschick, V.P., BassiriRad, H. 2003. Extreme events as shaping physiology, ecology, and evolution of plants: toward a unified definition and evaluation of their consequences. New Phytologist 160: 21-42.

Hacke, U.G., Sperry, J.S., Pittermann, J. 2000. Drought experience and cavitation resistance in six shrubs from the Great Basin, Utah. Basic and Applied Ecology 1: 31-41.

Hampe, A. 2004. Bioclimate envelope models: what they detect and what they hide. Global Ecology and Biogeography 13: 469-471.

Hanley, J.A., McNeil, B.J. 1982. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. Radiology 143: 29-36.

Hannah, L., Midgley, G.F., Millar, D. 2002. Climate change-integrated conservation strategies. Global Ecology and Biogeography 11: 485-495.

Hastie, T., Tibshirani, R. 2001. The Elements of Statistical Learning. Springer, New York.

Heikkinen, R.K., Luoto, L., Araújo, M.B., Virkkala, R., Thuiller, W., Sykes, M.T. 2006. Methods and uncertainties in bioclimatic envelope modeling under climate change. Progress in Physical Geography 30: 751-777.

Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29: 773-785.

Hijmans, R.J., Garrett, K.A., Huaman, Z., Zhang, D.P., Schreuder, N., Bonierbale, N., 2000. Assessing the geographic representativeness of genebank collections: the case of the Bolivian wild potatoes. Conservation Biology 14: 1755-1765.

Hiserote, B., Waddell, K., 2003. The PNW-FIA integrated database user guide. A database of forest inventory information for California, Oregon and Washington, version 2.0. FIA PNW Station, Portland, Oregon.

Holling, C.S. 1973. Resilience and Stability in Ecosystems. Annual Review of Ecology and Systematics 4: 1-23.

Holling, C.S. 1992. Cross-scale morphology, geometry, and dynamics of ecosystems. Ecological Monographs 62: 447-502.

Holmgren, C.A., Norris, J., Betancourt, J.L. 2006. Inferences about winter temperatures and summer rains from the late Quaternary record of C4 perennial grasses and C3 desert shrubs in the northern Chihuahuan Desert. Journal of Quaternary Science 22: 141-161.

Hugall, A., Moritz, C., Moussalli, A., Stanisic, J. 2002. Reconciling paleodistribution models and comparative phylogeography in the wet tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). Proceedings of the National Academy of Science, U.S.A. 99: 6112-6117.

Hutchinson, G.E. 1957. Concluding remarks. Cold Spring Harbor Sympiosia on Quantitiative Biology 22: 415-427.

Hutchinson, M.F., Gessler, F.R.1994. Splines: More than just a smooth interpolator. Geoderma 62: 45-67.

Ibanez, I., Clark, J.S., Dietze, M.C. 2009. Estimating the migration potential of migrant tree species. Global Change Biology 15: 1173-1188.

IPCC. 2001. Climate Change 2001, The Scientific basis. Cambridge University Press, New York.

Iverson, L.R., Prasad, A.M. 1998. Predicting abundance of 80 tree species following climate change in the eastern United States. Ecological Monographs 68: 465-485.

Jackson, S.T., Betancourt, J.L., Booth, R.K., Gray, S.T. 2009. Ecology and the ratchet of events: climate variability, niche dimensions, and species distributions. Proceedings of the National Academy of Sciences, U.S.A. 106: 19685-19692.

Jacquez, G.M., Maruca, S., Fortin, M.J. 2000. From fields to objects: A review of geographic boundary analysis. Journal of Geographical Systems 2: 221-241.

Jarvis, A., Williams, K., Williams, D., Guarino, L., Caballero, P.J., Mottram, G. 2005. Use of GIS for optimizing a collecting mission for a rare wild pepper (*Capsicum flexuosum* Sendtn.) in Paraguay. Genetic Resources and Crop Evolution 52: 671-682.

Jentsch, A., Kreyling, J., Beierkuhnlein, C. 2007. A new generation of climate change experiments: events, not trends. Frontiers in Ecology and the Environment 5: 315-324.

Jørgensen, S.E., Patten, B.C., Strakraba, M. 1992. Ecosystems emerging: toward an ecology of complex systems in a complex future. Ecological Modelling 62: 1-27.

Kadmon, R., Farber, O., Danin, A. 2003. A Systematic analysis of factors affecting the performance of climatic envelope models. Ecological Applications 13: 853-867.

Kadmon, R., Farber, O., Danin, A. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecological Applications 14: 401-404.

Kelly, A.E., Guolden, M.L. 2008. Rapid shifts in plant distribution with recent climate change. Proceedings of the National Academy of Sciences, United States of America 105:11823-11826.

Kelly, C.K., Bowler, M.G., Pybus, O.G., Harvey, P.H. 2008. Phylogeny, niches and relative abundance in natural communities. Ecology 89: 962-970.

Kerr, J.T., Kharouba, H.M., Currie, D.J. 2007. The macroecological contribution of global change solutions. Science 316:1581-1584.

Kimball, S., Angert, A.L., Huxman, T.E., Venable, D.L. 2010. Contemporary climate change in the Sonoran Desert favors cold-adapted species. Global Change Biology 16:1555-1565.

Kinzig, A.P., Ryan, P., Etienne, M., Allison, H., Elmqvist, T., Walker, B.H. 2006. Resilience and regime shifts: assessing cascading effects. Ecology and Society 11: 20-36.

Kozlowski, T.T., Kramer, P.J., Pallardy, S.J. 1991. The Physiological Ecology of Woody Plants. Academic Press, New York.

Landsberg, J.J., Waring, R.H. 1997. A Generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance, and partitioning. Forest Ecology and Management 95: 209-228.

Larcher, W. 2003. Physiological Plant Ecology. Springer, Germany.

Leibold, M. A. 1995. The niche concept revisited: mechanistic models and community context. Ecology 76: 1371-1382.

Lenoir, J., Gégout, J.C., Pierrat, J.C., Bontemps, J.D, Dhôte, J.F. 2009. Differences between tree species seedling and adult altitudinal distribution in mountain forests during the recent warm period (1986-2006). Ecography 32: 765-777.

Levin S. 1999. Fragile Dominion: Complexity and the Commons. Perseus Books, Massachusetts.

Levitt, J. 1980. Response of Plants to Environmental Stresses, Vol. 1, 2nd ed. Academic Press, New York.

Limburg, K.E., O'Neill, R., Costanza, R., Farber, S. 2002. Complex systems and valuation. Ecological Economics 41: 409-420.

Lin, X., Carroll, R.J., 2000. Non-parametric function estimation for clustered data when the predictor is measured without/with error. Journal of the American Statistical Association 95: 520-534.

Linton, M.J., Sperry, J.S., Williams, D.G. 1998. Limits to water transport in *Juniperus osteosperma* and *Pinus edulis*: implications for drought tolerance and regulation of transpiration. Functional Ecology 12: 906-911.

Lintz, H., McCune, B., Gray, A., McCulloh, K. 2010. Quantifying ecological thresholds from response surfaces. Ecological Modelling, In press.

Loehle, C. 1998. Height growth rate trade offs determine northern and southern range limits for trees. Journal of Biogeography 25: 735-742.

Lyons, J.M., Graham, D., Raison, J.K. 1979. Low Temperature Stress in Crop Plants: the Role of the Membrane. Academic Press, New York.

Macarthur, R. 1972. Geographical ecology: patterns in the distribution of species. Harper and Row, New York.

Makarewicz, J.C., Likens, G.E. 1975. Niche analysis of a zooplankton community. Science 190: 1000-1003.

Mantua, N.J., Hare, S.R., Zhang, Y., Wallace, J.M., Francis, R.C. 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. Bulletin of the American Meteorological Society 78: 1069-1079.

Marini, M.A., Barbet-Massin, M., Lopes, L.E., Jiguet, F. 2009. Predicted climate-driven bird distribution changes and forecasted conservation conflicts in a neotropical savanna. Conservation Biology 23: 1558-1567.

Massey, F.J. Jr. 1951. The Kolmogorov-Smirnov test of goodness of fit. Journal of the American Statistical Association 46: 68-78.

Maurer, B.A. 1998. Ecological science and statistical paradigms: at the threshold. Science 279: 502.

McCabe, G.J., and Dettinger, M.D. 1999. Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States. International Journal of Climatology 19:1399-1410.

McCabe, G.J., Betancourt, J.L., Gray, S.T., Palecki, M.A., Hidalgo, H.G. 2008. Association of multi-decadal sea surface temperature variability with U.S. drought. Quanternary International 188: 31-40.

McCune, B., Keon, D. 2002. Equations for potential annual direct incident radiation and heat load. Journal of Vegetation Science 13: 603-606.

McCune, B. 2006. Non-parametric habitat models with automatic interactions. Journal of Vegetation Science 17: 819-830.

McCune, B., Mefford, M.J. 2006. PC-ORD. Multivariate analysis of ecological data, Version 5.2. MjM Software Design, Gleneden Beach, Oregon.

McCune, B., Mefford, M.J. 2008. Hyperniche. Multiplicative Habitat Modeling. Version 2.0. MjM Software, Gleneden Beach, Oregon.

McDowell, N., Pockman, W.T., Allen, C.G., Breshears, D.D., Cobb, N., Kolb, T., Plaut, J., Sperry, J., West. A., Williams, D.G., Yepez, E.A. 2008. Mechanisms of plant survival and mortality during drought: why do some plants survive while others succumb to drought? New Phytologist 178: 719-739.

McKenzie, D., Peterson, D.W., Peterson, D.L., Thornton., P.E. 2003. Climatic and biophysical controls on conifer species distributions in mountain forests of Washington State, U.S.A. Journal of Biogeography 30: 1093-1110.

Mielke, P.W., Jr., Berry, K.J. 2001. Permutation Methods: A Distance Function Approach. Springer Series in Statistics. Springer Science, New York.

Miller, J., Franklin, J., Aspinall, R. 2007. Incorporating spatial dependence in predictive vegetation models. Ecological Modelling 202: 225-242.

Murray, F.W. 1967. On the computation of saturation vapor pressure. Journal of Applied Meteorology 6: 203-204.

Myneni, R.B., Kelling, C.D., Tucker, C.J., Asrar, G., Nemani, R.R. 1997. Increased plant growth in the northern high latitudes from 1981 to 1991. Nature 386: 698-702.

National Research Council. 2000. Ecological Indicators for the Nation. National Academy Press, Washington D.C.

Nelson, B.W., Ferreira, C.A.C., da Silva, M.F., Kawasaki, M.L. 1990. Endemism centres, refugia and botanical collection density in Brazilian Amazonia. Nature 345: 714-716.

Neilson, R.P., Wullstein, L.H. 1983. Biogeography of two southwest American oaks in relation to atmospheric dynamics. Journal of Biogeography 10: 275-297.

Niering, W.A., Whittaker, R.H., Lowe, C.H. 1963. The saguaro, a population in relation to eviroment. Science 142: 15-23.

Newbold, S., Eadie, J.M. 2004. Using species-habitat models to target conservation: a case study with breeding mallards. Ecological Applications 14: 1384-1393.

Odum, E. 1971. Fundamentals of Ecology, Third Edition. W.B. Saunders Company, Philadelphia.

Ohmann, J.L., Spies, T.A. 1998. Regional gradient analysis and spatial pattern of woody plant communities of Oregon forests. Ecological Monographs 68: 151-182.

Ohmann, J.L., Gregory, M.J. 2002. Predictive mapping of forest composition and structure with direct gradient analysis and nearest neighbor imputation in coastal Oregon, U.S.A. Canadian Journal of Forest Research 32: 725-741.

Oksanen, J., Minchin, P.R. 2002. Continuum theory revisited: what shape are species' responses along ecological gradients? Ecological Modelling 157: 119-129.

Oren, R., Sperry, J.S., Katul1, G.G., Pataki, D.E., Ewers, B.E., Phillips, N., Schafer, K.V.R. 2002. Survey and synthesis of intra- and interspecific variation in stomatal sensitivity to vapour pressure deficit. Plant, Cell, and the Environment 22: 1515-1526.

Parmesan, C., Root, T.L., Willig, M.R. 2000. Impacts of extreme weather and climate on terrestrial biota. Bulletin of American Meteorological Society 81: 443-50.

Peñuelas, J., Filella, I., Zhang, X., Llorens, L., Ogaya, R., Lloret, F., Comas, P., Estiarte, M., Terradas, J. 2004. Complex spatiotemporal phenological shifts as a response to rainfall changes. New Phytologist 161: 837-846.

Pearce, J., Ferrier, S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling 133: 225-245.

Peterson, A.T., Soberón, J., Sanchez-Cordero, V. 1999. Conservatism of ecological niches in evolutionary time. Science 285: 1265-1267.

Peterson, A.T., Ortega-Huerta, M.A., Bartley, J., Sánchez-Cordero, V., Soberón, J., Buddemeier, R.H., Stockwell, D.R.B. 2002. Future projections for Mexican faunas under global climate change scenarios. Nature 416: 626-629.

Peterson, A.T. 2003. Predicting the geography of species' invasions via ecological niche modeling. Quarterly Review of Biology 78: 419-433.

Raffa, K.F., Aukema, B.H., Bentz, B.J., Carroll, A.L., Hicke, J.A., Turner, M.G., Romme, W.H. 2008. Cross-scale drivers of natural disturbances prone to anthropogenic amplification: The dynamics of bark beetle eruptions. BioScience 58: 501-517.

Raffa, K.F., Aukema, B.H., Erbilgin, N., Klepzig, K.D., Wallin, K.F. 2005. Interactions among conifer terpenoids and bark beetles across multiple levels of scale: An attempt to understand links between population patterns and physiological processes. Recent Advances in Phytochemistry 39: 80-118.

Raison, J.K., Chapman, E.A., Wright, I.C., Jacobs, S.W.L. 1979. Membrane lipid transitions: their correlation with the climatic distributions of plants. Pg. 177-186.

IN Lyons, J.M., Graham, D., Raison J.K. 1979. Low Temperature Stress in Crop Plants: the Role of the Membrane. Academic Press, New York.

Raxworthy, C.J., Martinez-Meyer, E., Horning, E. Nussbaum, R.A., Schneider, G.E., Ortega-Huerta, M.A., Townsend-Peterson, A. 2003. Predicting distributions of known and unknown reptile species in Madagascar. Nature 426: 837-841.

Reese, G.C., Wilson, K.R., Hoeting, J.A., Flather, C.H. 2005. Factors affecting species distribution predictions: a simulation modeling experiment. Ecological Applications 15: 554-564.

Rehfeldt, G.E., Crookston, N.L., Warwell, M.., Evans, J.S. 2006. Empirical analyses of plant-climate relationships for the western United States. Internaitonal Journal of Plant Science 167: 1123-1150.

Rehfeldt, G.E., Ferguson, D.E., Crookston, N.L. 2008. Quantifying the abundance of co-occurring conifers along the inland northwest (USA) climate gradients. Ecology 89: 2127-2139.

Ropelewski, C., Halpert, M. 1986. North American precipitation and temperature patterns associated with the El Nino/Southern Oscillation (ENSO). Monthly Weather Reviews 114: 2352-2362.

Ruckstuhl, A., Welsh, A.H., and Carroll, R.J. 2000. Nonparametric function estimation of the relationship between two repeatedly measured variables. Statistica Sinica 10: 51-71.

Rudis, V.A. 2003a. Comprehensive regional resource assessments and multipurpose uses of Forest Inventory and Analysis data, 1976 to 2001: A Review. General Technical Report SRS-70. U.S. Department of Agriculture, Forest Service, Southern Research Station. Ashville, North Carolina

Rudis, V.A. 2003b. Fresh ideas, perspectives, and protocols associated with Forest Inventory and Analysis surveys: graduate reports, 1974 to July 2001. General Technical Report SRS-61. U.S. Department of Agriculture, Forest Service, Southern Research Station. Ashville, North Carolina

Schabenberger, O, Gotway, C.A., 2005. Statistical Methods for Spatial Data Analysis. Chapman & Hall/CRC, Boca Raton, Florida.

Scheffer, M., Carpenter, S.R. 2003. Catastrophic regime shifts in ecosystems: linking theory to observation. Trends in Ecology and Evolution 18: 648-656.

Scheffer, M., Hosper, S.H., Meijer, M.L., Moss, B., Jeppesen, E. 1993. Alternative equilibria in shallow lakes. Trends in Ecology and Evolution 8: 275-279.

Schlesinger, M.E., Ramankutty, N. 1994. An oscillation in the global climate system of period 65-70 years. Nature 367: 723-726.

Schrag, A.M., Bunn, A.G., Graumlich, L.J. 2008. Influence of bioclimatic variables on tree line conifer distribution in the Greater Yellowstone Ecosystem: Implications for species of conservation concern. Journal of Biogeography 35: 698-710.

Scott, C.T., Cassell, D.L., Hazard, J.W. 1993. Sampling design of the U.S. National Forest Health Monitoring Program. In: Nyyssonen, A., Poso, S., Rautala, J. (eds.), Ilvessalo symposium on National Forest Inventories. Finnish Forest Research Institute Research Papers 444: 150-157.

Segurado, P., Araújo, M.B., Kunin, W.E. 2006. Consequences of spatial autocorrelation for niche-based models. Journal of Applied Ecology 43: 433-444.

Shafer, S.L., Bartlein, P.J., Thompson, R.S. 2001. Potential changes in the distributions of western North America tree and shrub taxa under future climate scenarios. Ecosystems 4: 200-215.

Shreve, F. 1991. The influence of low temperature on the distribution of the giant cactus. Plant World 14: 136-146.

Smith, W.B., Miles, P.D., Vissage, J.S., Pugh. P.A. 2004. Forest Resources of the United States. North Central Research Station Forest Service, U.S. Department of Agriculture. St. Paul, Minnesota

Soberón, J., Nakamura, M. 2009. Niches and distributional areas: concepts, methods, and assumptions. Proceedings of the national Academy of Sciences, U.S.A. 106: 19644-19650.

Sperry, J.S., Donnelly, J.R., Tyree, M.T. 1988. A method for measuring hydraulic conductivity and embolism in xylem. Plant Cell and the Environment 11: 35-40.

Sperry, J.S., Sullivan, J.E.M. 1992. Xylem embolism in response to freeze-thaw cycles and water stress in ring-porous, diffuse-porous, and conifer species. Plant Physiology 100: 605-613.

Sperry, J.S., Meinzer, F.C., McCulloh, K.A. 2008. Safety and efficiency conflicts in hydraulic architecture: scaling from tissues to trees. Plant Cell and the Environment 31: 632-645.

Stenseth, N.C., Mysterud, A., Ottersen, G., Hurrel, J.W., Chan, K., Lima, M. 2002. Ecological effects of climate fluctuations. Science 297: 1292-1296.

Stockwell, D.R.B., Peterson, A.T. 2002. Effects of sample size on accuracy of species distribution models. Ecological Modelling 148: 1-13.

Sutton, R.T., Hodson, D.L.R. 2005. Atlantic Ocean forcing of North American and European summer climate. Science 309: 115-118.

Svenning, J., Skov, F. 2004. Limited filling of the potential range in European tree species. Ecology Letters 7: 565-573.

Swenson, J., Waring, R.H. 2006. Modelled photosynthesis predicts woody plant richness at three geographic scales across the North-Western United States. Global Ecology and Biogeography 15: 470–485.

Swetnam, T.W., Betancourt, J.L. 1998. Mesoscale disturbance and ecological response to decadal climatic variability in the American Southwest. Journal of Climatology 11: 3128-3147.

Swetnam, T.W., Allan, C.D., Betancourt, J.L. 1999. Applied historical ecology: using the past to manage for the future. Ecology 9: 1189-1206.

Tanaka, H., Shibata, M., Masaki, T., Iida, S., Niiyama, K., Abe, S., Kominami, Y., Nakashizuka, T. 2008. Comparative demography of three coexisting Acer species in gaps and under closed canopy. Journal of Vegetation Science 19: 127-138.

Taylor, P. 2005. Unruly complexity. University of Chicago Press, Chicago.

Thom, R. 1989. Structural Stability and Morphogenesis:An Outline of a General Theory of Models. Addison-Wesley, Massachussetts.

Thornton, P.E., S.W. Running, White, M.A. 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. Journal of Hydrology 190: 214-251.

Thuiller, W., Brotons, L., Araújo, M.B., Lavorel, S. 2004. Effects of restricting environmental range of data to project current and future species distributions. Ecography 27: 165-172.

Toms, J.D., Lesperance, M.L. 2003. Piecewise regression: a tool for identifying ecological thresholds. Ecology 84: 2034-2041.

Trenberth, K. 1984. Signal versus noise in the Southern Oscillation. Monthly Weather Review 112: 326-332.

Trenberth, K., Hoar, T.J. 1996. The 1990-1995 El Niño-Southern Oscillation event: the Longest on record. Geophysical Research Letters 23: 57-60.

Trenberth, K. 1997. The Definition of El Niño. Bulletin of the American Meterological Society 78: 2771-2777.

Trenberth, K. E., Stepaniak, D.P. 2001. Indices of El Niño evolution. Journal of Climate 14: 1697-1701.

U. S. Climate Change Science Program and the Subcommittee on Global Change Research. 2009. Thresholds of climate change in ecosystems. USCCP Publication, Virginia.

U.S.D.A. Forest Service. 1992. Forest service resource inventories: an Overview. Washington, D.C.

von Hayek, F.A.1964. "The Theory of Complex Phenomena" in The Critical Approach to Science and Philosophy. Essays in Honor of K.R. Popper, ed. M. Bunge, New York.

van Mantgem, P.J., Stephenson, N.L., Byrne, J.C., Daniels, L.D., Franklin, J.F., Fulé,P.Z., Harmon, M.E., Larson, A.J., Smith, J.M., Taylor, A.H., Veblen, T. 2009. Widespread increase of tree mortality rates in the western United States. Science 323: 521-524.

Vogt, U.K. 2001. Hydraulic vulnerability, vessel refilling, and seasonal courses of stem water potential of *Sorbus aucuparia* L. and *Sambucus nigra* L. Journal of Experimental Botany 52: 1527-1536.

Wahba, G. 1990. Comment on Cressie. The American Statistician 44: 255-256.

Walsh, P.D., Lawler, D.M. 1981. Rainfall seasonality: description, spatial patterns and change through time. Weather 36: 201-208.

Walther, G.R., Convey, P., Post, E., Menzel, A., Parmesan, C., Beebee, T.J., Fromentin, J., Hoegh-Guldberg, O., Bairlein, F. 2002. Ecological responses to recent climate change. Nature 416: 389-395.

Waring, R.H., Major, J. 1964. Some Vegetation of the California coastal redwood region in relation to gradients of moisture, nutrients, light, and temperature. Ecological Monographs 34: 167-215.

Weltzin, J.F., Loik, M.E., Schwinning, S., Williams, D.G., Fay, P., Haddad, B., Harte, J., Huxman, T.E., Knapp, A.K., Lin, G., Pockman, W.T., Shaw, M.R., Small, E., Smith, M.D., Smith, S.D., Tissue, D.T., Zak, J.C. 2003. Assessing the response of terrestrial ecosystems to potential changes in precipitation. Bioscience 53: 941-952.

Weng, G., Bhalla, U.S., Iyengar, R. 1999. Complexity in biological signaling systems. Science 284: 92-96.

Wessels, K.J., Van Jaarsveld, A.S., Grimbeek, J.D.,Van der Linde, M.J., 1998. An evaluation of the gradsect biological survey method. Biodiversity Conservation 7: 1093-1121.

Whittaker, R.H. 1975. Communities and Ecosystems, 2nd Edition. Macmillan, New York.

Wikle, C.K. 2003. Spatio-temporal models in climatology. In Encyclopedia of Life Support Systems (EOLSS), Developed under the Auspices of the UNESCO, Eolss Publishers, Oxford.

Wilcoxon, F. 1945. Individual comparisons by ranking methods. Biometrics 1: 80-83.

Wilk, M.B., Gnanadesikan, R. 1968. Probability plotting methods for the analysis of data. Biometrika Trust. 55: 1-17.

Woodall, C.W., Oswalt, C.M., Westfall, J.A., Perry, C.H., Nelson, M.D., 2009a. Tree migration detection through comparisons of historic and current forest inventories. In: McWilliams, W., Moisen, G., Czaplewski, R., (eds.), Forest Inventory and Analysis Symposium 2008, Park City, UT. Proc. RMRS-P-56CD. Fort Collins, Colorado.

Woodall, C.W., Oswalt, C.M., Westfall, J.A., Perry, C.H., Nelson, M.D., Finley, A.O., 2009b. Selecting tree species for testing climate change migration hypotheses using forest inventory data. Forest Ecology and Management 259: 778-785.

Womble, W.H. 1951. Differential systematics. Science 114: 315-322.

Woodward, F.I. 1987. Climate and plant distribution. Cambridge University Press. Cambridge, United Kingdom.

Yakowitz, S. J., Szidarovszky, F. 1985. A comparison of kriging with nonparametric regression methods. Journal of Mulitvariate Analysis 16: 31-53.

Yanez, M., Floater, G. 2000.Spatial distribution and habitat preference of the endangered tarantula *Brachypelma klassi* (Araneae: Theraphosidae) in Mexico. Biodiversity and Conservation 9: 795-810.

Yost, A. 2008. Probabilistic modeling and mapping of plant indicator species in a Northeast Oregon industrial forest, U.S.A. Ecological Indicators 8: 46-56.

Zhang, Y., Wallace, J.M., Battisti, D.S. 1997. ENSO-like interdecadal variability: 1900-1993. Journal of Climate 10: 1004-1020.

Zhang, R., Delworth,T. L. 2007. Impact of the Atlantic Multidecadal Oscillation on North Pacific climate variability. Geophysical Research Letters 34: L23708.

Zhao, M., Running, S.W. 2010. Drought-induced reduction in global terrestrial net primary production from 2000 Through 2009. Science 329: 940-943.

Zobel, D.B. 1974. Local variation in intergrading *Abies grandis-A. concolor* populations in the central Oregon Cascades. II. Stomatal reaction to moisture stress. Botatical Gazette 135: 200-210.

Zon, R. 1904. Effects of frost upon forest vegetation. Forestry Quarterly 2: 14-21.